Intelligent Automation And Soft Computing, 2020
Vol. 26, no. 3, 585–595
DOI: 10.32604/iasc.2020.013937

The Factor Analysis of University English Examination Results Based on the Multilevel Model

Shaoyun Long[1,*], Qianying Long[2]

[1]Foreign languages college, Jiangxi Normal University Jiangxi, China 330022
[2]College of Tourism and Service Management, Nankai University Tianjin, China 300350
Address: no. 99, Ziyang avenue, Nanchang City, Jiangxi province

ABSTRACT
The traditional factor analysis models, such as the generalized linear regression model and the gistic regression model have disadvantages of large standard error of analysis results. For this purpose, a multilevel factor analysis model based on time series and independent variable data is designed. The OLS estimation analysis method is used to establish the basic environment form, to derive the model calculation parameters and to complete the environment construction of the multilevel analysis model. On the basis of the construction environment, the double-level environment reference module and the multilevel factor analysis module are designed to realize the design of the multi water leveling factor analysis model. Compared with the traditional factor analysis model, the standard error of the analysis results are reduced by 15%.

KEY WORDS: Influencing Factors; level model, University English achievement, standard error.

1 INTRODUCTION

THE multilevel model is applied to the analysis of hierarchical data, which was proposed by Lindley and Smith in 1972 but because of the limitations of the traditional parameter estimation methods (ordinary least squares estimation) and computing technology, the multilevel model has not been extended and applied (LengQingfan, 2016). Then, other scholars such as Dempste. R, Laird and Rubin improved the model parameter estimation, and put forward the EM algorithm, which proposed the iterative weighted generalized least square method and the Fisher scoring algorithm, thus the development of corresponding software, such as HIM, Mlwin and VARCL, promoted the development of the theory and application of the multilevel model. In abroad, multilevel models are widely used in social sciences, basic medicine, clinical medicine, preventive medicine and other fields (Dai Jia, 2017). It is used in the field of psychology and the analysis of examination results but is mainly based on the data of continuous variables, and it has not been seen for the analysis of the data of classification variables (such as the passing rate and the excellent rate).

The score of theCET-4 is an important index to measure the level of university English teaching and the degree of English mastery of university students (Chen Jinyang, 2017). The examination is authoritative, strict and standard, and the results are comparable. In this study, based on the determination of the multilevel analysis environment, the examination result factor analysis model is established. The model is used to analyze the influence factors of the CET-4 examination results of 20 college students of the 2003 level in a university, and compare the calculation results of the traditional algorithm with the results of the proposed algorithm in this study (Liu Miao, 2017). It confirms the applicability of the multilevel model in the application of examination scores analysis and completes the research in this study.

2 DETERMINATION OF THE ENVIRONMENT ESTABLISHED BY THE MULTILEVEL ANALYSIS MODEL

2.1 *The Determination Method*

MOST linear analysis relies on the ordinary least square estimation (OLS) (Liu Jiezhen, 2017), and the multilevel model uses shrinkage estimation, which is more stable or accurate than the "regression" by using OLS (Wang Lanlan, 2017). If some second level units have only a small number of individual samples, the

CONTACT Shaoyun Long sgdxg0@163.com

multilevel model uses the weighted sum of the two estimates as the final estimate. One is the OLS estimation from each double-level unit and the other is the weighted least squares (WLS) estimation between the second levels. The final estimate is weighed by the sample size of the working group. The smaller size of the sample is more dependent on the second layer, the WLS estimation, and the larger size of the sample wood depends on the first layer of the OLS estimate (Guo Dong, 2017).

2.2 Determination of the Construction of the Basic Environment Form

According to research purposes and information of the researchers(Hu Jianxiang, 2017), the multilevel model can have different forms, such as; double-classification discrete data multilevel model, multi-discrete data multilevel model, multilevel model of repeated measurement data, multilevel cross classification model, bivariate multilevel model, nonlinear multilevel model, multilevel time series model and so on. But its basic form consists of 3 equations(Huang Shafei, 2017):

$$Y_{\text{ij}} = \beta_{oj} + \beta_{ij} \times X_{\text{ij}} + \text{R}_{\text{ij}}$$

$$\beta_{oj} = \gamma_{oo} + \mu_{oj}$$

$$\beta_{\text{lj}} = \gamma_{lo} + \mu_{lj} \quad (1)$$

where i represents the first layer of individuals such as students and patients; j represents the second layer of units belonging to the first layer, such as schools, and classes γ_{oo} represents the mean value of β_{oj} ; γ_{lo} represents the mean value of β_{lj} , and they are constant between units of the second layer. They are fixed components of β_{oj} and β_{lj} ; μ_{oj} represents the random formation of $\beta_{\text{o j}}$. μ_{lj} represents the random component of β_{lj} , which represents the variation between second levels of units (Sun Xiaojian, 2016).

The variance and covariance are:

$$Var(\mu_{\text{oj}}) = \text{T}_{oo}$$

$$Var(\mu_{lj}) = T_{ll}$$

$$Cov(\mu_{oj}, \mu_{l}) = T_{\text{ol}} \quad (2)$$

Replace the corresponding terms of $Y_{\text{ij}} = \beta_{\text{oj}} + \beta_{lj} \times X_{lj} + R_{\text{ij}}$ in the equation by using $\beta_{\text{oj}} = \gamma_{oo} + \mu_{oj}$ and $\beta_{\text{lj}} = \gamma_{lo} + \mu_{lj}$ in Equation (1), then:

$$Y_{\text{ij}} = \gamma_{\text{oo}} + \gamma_{lo} \times X_{\text{ij}} + \mu_{\text{oj}} + \mu_{\text{lj}} \times X_{\text{ij}} + R_{ij} \quad (3)$$

In Equation (3), $(\mu_{\text{oj}} + \mu_{\text{lj}} \times X_{\text{ij}} + R_{\text{ij}})$ is a residual term. Since each individual in second layers has the same μ_{oj} and μ_{lj}, the similarity between the individuals within the same second layer units is higher than that of the individuals from units of different second layers, which is the source of the related residual (Yu Xuejiao, 2015). Because the values of μ_{oj} and μ_{lj} are different, the residuals from the second levels of the different units may have different variances. The error items are related, the variance is not equal, and are related to X_{lj} . If there is no difference between the second level units, the value of (Cheng Yong, 2015), μ_{oj} and μ_{lj} are 0, and equation (3) is simplified to a simple OLS regression, that is, $Y_{\text{i}} = \beta_{o} + \beta_{i} \times X_{\text{i}} + R_{i}$. When data exist in a hierarchical structure, the value of μ_{oj} and μ_{lj} are not 0, and cannot be analyzed by the simple OLS regression. Instead, the OLS should be combined with the multilevel models (Sun Lin, 2015).

2.3 Determination of the Calculation Parameters of the Model

2.3.1 The Selection of Time Series

It is assumed that in the double-level model of the repeated measurement, level 1 residual e_{oij} obeys the first order autoregressive process, and the time point of the measurement is of the equal distance(Ding Cheng, 2015), then the discrete time calculation equation at time t can be derived.

$$Y_{ij(1)} = \sum_{h=0}^{P} U_{hXhij} + \sum_{h=0}^{Q} U_{\text{hjZhij}} + (D\text{e}_{\text{oij(t-1)}} + Vt) \quad (4)$$

In this equation, $\text{e}_{oij(t-1)}$ represents the horizontal residuals of t-1 and D is the autoregressive coefficient between 1 residual e_{oij} of adjacent time.

When the observation time is not equal, the continuous time series equation can be described and the subscript of level 1 and level 2 is omitted, and the general equation of the horizontal 1 residual can be written as (Yu Feifei, 2016):

$$COV(\text{e}_{\text{t}} e_{t-s}) = e_{e}^{2} f(s) \quad (5)$$

In this way, the covariance between the two measurements depends on a variance function (HaoGuang, 2016), which also depends on the time of each measurement, the time difference in the measurement, and the other time difference function. Among them, the specific form of f(s) can be determined according to the actual requirements. When multiple measurements are closely related to time, the continuous measurement may lead to the correlation trend of horizontal 1 residual, and the introduction of the time series into a multilevel model and can clearly identify the autocorrelation structure (Cui Junying, 2015). The characteristic of the multilevel time series model is to establish an autoregressive model for the horizontal residuals, which can reflect the effect of time by parameter D in estimated model (4). In the continuous time models, there are allowable missing values and different function forms for autoregressive coefficients, to facilitate data convergence. For a model, continuous time is more flexible than discrete time, but it is also more complex. Therefore, in terms of the application, it should be determined according to specific circumstances (Hu Liang, 2015).

The traditional time series equation only analyzes the change of data in time and predicts possible future trends, mainly used in the data of long term time series, and the introduction of time series in the multilevel model is the part of solving the repeated measurement of data. The data variation is mainly used for the short time series data (Lu Chenlong, 2015).

2.3.2 Selection of Discrete Data

First, two classifications of discrete data are determined. When the response variable is a double-value variable, the two-item distribution is generally used. When it is counted as a unit, Poisson distribution is usually used to fit it. Taking the double-value variable as an example, the Logitech connection function is used to fit the data parameter of the double-level model (Qin Yin-yi, 2015). The equation is as follows:

$$C_{\mathrm{ij}} = (1+\exp\ (-[\sum_{\mathrm{h}=0}^{\mathrm{P}} U_{hXhij} + \sum_{\mathrm{h}=0}^{Q} U_{hjZhij}])\ )^{-1} \tag{6}$$

C_{ij} is the expected value of the 1-unit response at level ij.

The response variable $Y_{ij} \sim B\mathrm{in}\ (C_{\mathrm{ij}},\ n_{ij})$ can be used to write equations including the horizontal 1 variation.

$$Y_{ij} = C_{\mathrm{ij}} + e_{ij}\sqrt{e_{ij}(1-e_{ij}) \div n_{ij}} \tag{7}$$

$e_{ij}\sqrt{e_{ij}(1-e_{ij}) \div n_{ij}}$ represents level 1 residuals, because Y_{ij} obeys the two-item distribution, so e_{ij} is corrected so that residuals are subject to normal distribution.

Secondly, the multi classification discrete data is determined by;

(1) When the data is in a disordered classification, assuming that the reaction variables have t classification, they are considered as t response variables, and select one of (t) as the base class. The remaining t-1 reaction variables are calculated according to equation (6) respectively. Therefore, the Logit equation of the multivariate data of t-1 can be written as follows:

$$\log[C_{\mathrm{ij}}^{(s)} \div C_{ij}^{(\mathrm{t})}] = \sum_{\mathrm{h}=0}^{P} U_{hX\mathrm{hij}}^{(s)} + \sum_{h=0}^{Q} U_{\mathrm{hjZhij}}^{(s)} \qquad s = 1,2,3,\ldots,t-1 \tag{8}$$

And $\sum_{h=1}^{\mathrm{t}} e_{ij}^{(s)} = 1$.

(2) When the data is in an orderly classification, the cumulative variable rate equation for the ordered categorical data is:

$$V_{ij}^{(s)} = \sum_{h=1}^{s} C_{\mathrm{ij}}^{h} \qquad s = 1,2,3,\ldots,t-1 \tag{9}$$

By using the Logit connection function, the data can be obtained.

$$V_{ij}^{(\mathrm{s})} = \{1+\exp[-(\sum_{\mathrm{h}=0}^{\mathrm{P}} U_{hXhij}^{(s)} + \sum_{h=0}^{Q} U_{hjZ\mathrm{hij}}^{(s)})]\}^{-1} \tag{10}$$

By using the log-log connection function, the data proportional risk equation can be obtained as:

$$V_{ij}^{(s)} = \{1-\exp[-\exp(\sum_{h=0}^{P} U_{hXhij}^{(s)} + \sum_{h=0}^{Q} U_{\mathrm{hjZhij}}^{(s)})]\} \tag{11}$$

When the data is in a disordered classification, the covariance structure is complex, the iteration is slow and even difficult to converge, so the data classification should not be too much, the sequential data classification is convergent fast and the result is easy to explain. Sometimes the continuous response variable can be converted into an ordered classification reaction variable and is used to deal with the multi classification reaction variable equation.

In practice, different connection functions can be used according to different data, such as Logit, Probit and so on. In this study, the Logit connection function is used. Generally, the double-level discrete equations are widely used, while three and higher-level discrete equations are less discussed, because their structure is more complex. In practical applications, there are three ways to deal with three horizontal discrete equations. One is neglecting the one level and considering it as a fixed effect; two is the approximate process, that is, the model linearized by the first and two order Taylor series expansions of the logarithmic condition, and three is the full maximum edge estimation method. By using the digital integration to approximate multiple random applications, the three-level discrete model fitted is more accurate.

2.3.3 Determination of the Latent Variable

With L level and M1 latent variable at level 1, the equation of the multiple latent variables is:

$$Y_{ij} = \sum_{\mathrm{h}=0}^{P} U_{\mathrm{h}X\mathrm{hij}} + \sum_{l=2}^{L} \sum_{m=1}^{M1} u_m^{(l)} \lambda_m^{(l)} Z_{\mathrm{m}}^{(l)} \tag{12}$$

The first item on the right side of the equation is the fixed effect. The second item is a random effect, in which $\mathrm{u}_m^{(l)}$ represents the m latent variable in the horizontal 1, $Z_{\mathrm{m}}^{(\mathrm{l})}$ is the explanatory variable, $\lambda_m^{(l)}$ is the coefficient, and when $Z_m^{(l)} = Z_1^{(l)}$, there is a random intercept.

The multilevel latent variable equation can also be a special case of the GLLAMMs (generalized linear latent and mixed models), which also includes a multilevel factor equation, a generalized multilevel structural equation model, and so on. Its characteristic is that the equation contains a latent variable (latent variable), and a solution is used. The variable can be interpreted as a random effect (random intercept and coefficient), and the latent variable can also be changed at different levels, so there are random effects of level 2, level 3 and so on. The type of latent variable can be the continuous variable, discrete variable or multivariate normal distribution variable.

To sum it up, based on determining the estimation method, the basic form and calculation parameters of the model, the environment of the multilevel factor analysis model is established. After setting up the environment, a multilevel examination factor analysis model will be built based on the environment.

3 THE ESTABLISHMENT OF THE FACTOR ANALYSIS MODEL FOR EXAMINATION RESULTS BASED ON THE MULTI LEVEL ENVIRONMENT

AFTER determining the multilevel model analysis environment, the multilevel model can be used to decompose the random error to the different levels of the hierarchical data. The corresponding model, such as the double-level model or the triple-level model can be constructed, and the most common one is the two-horizontal model. This study will take the double-level model as an example to build an analysis model of university English examination results based on the multilevel environment. In the construction of the model, the first level is the student, the second level is the university. The first level equation is similar to the linear regression equation but its intercept and slope is the random variable; the intercept of the first level equation is the intercept, and th eslope depends on the second level (university) independent variable, thus forming a double-level model.

According to the variable types of the dependent variable, the corresponding double-level model is constructed, and the 2 double-level models are combined into a multilevel model. The process of building a model is described as follows:

3.1 Construction of the Double-level Environmental Reference Module

First, we construct a double-level model with numerical variables as dependent variables.

The first step is to establish the basic form of the model. First when constructing the model, the calculation equation of the first level is set as:

$$Y_{ij} = \beta_{0j} + \beta_{1j} \times X_{1\mathrm{ij}} + R_{ij} \tag{13}$$

The calculation equation of the second level is set as:

$$\beta_{0\mathrm{j}} = \gamma_{00} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10} + \mu_{1\mathrm{j}} \tag{14}$$

In the equation, subscript i represents the serial number of the first level unit (university student), i=1, 2, 3, ..., N, in which N is the sample content; the subscript j represents the number of the second level units (university students) subordinate to the first level, j=1, 2, 3,..., J, of which J is a second level unit.

Equation (13) is the first level of the model, in which Y_{ij} is the value of the dependent variable of the i student of the j university; $X_{1\mathrm{ij}}$ is the value of the first independent variable of the first j university student I; $\beta_{0\mathrm{j}}$ is the intercept, that is, when all the independent variables are 0, the mean value of the variable; $\beta_{1\mathrm{j}}$ is the regression coefficient of the first independent variable X_{1ij}. If there are more independent variables in the first level, then the independent variable is X_{2ij}, $X_{3\mathrm{ij}}$ ..., the corresponding partial regression coefficients are $\beta_{2\mathrm{j}}$, $\beta_{3\mathrm{j}}$... And R_{ij} is a random effect, which represents the partial, which cannot be explained by the independent variable in variable Y_{ij}.

Equation (14) is the second level of the model with β_{0j} and β_{1j} as the dependent variables. γ_{00} and γ_{10} in the equation are the mean values of β_{0j} and β_{1j} respectively. When the second level units are fixed, it is the fixed effect of β_{0j} and β_{1j}; μ_{0j} and μ_{1j} are the random effects of β_{0j} and β_{1j}, and are constantly changing within the second flat units.

The second step is to set the equation for the variance and covariance according to the random effect in the equation.

$$Var(R_{ij}) = \sigma^2$$
$$Var(\mu_{0j}) = \tau_{00}$$
$$Var(\mu_{1j}) = \tau_{1i}$$
$$Cov（\mu_{0j}, \ \mu_{1j}) = \tau_{01} \tag{15}$$

Among them, the variance of the first level random effect is σ^2, the variance of the second level random effect is τ_{00} and τ_{1i}, and the covariance is τ_{01}.

The third step is to establish a double-level model with numerical variables as dependent variables. (1) Set up the unconditional equation and the unconditional equation, also called variance component analysis equation:

$$Y_{ij} = \beta_{0j} + R_{ij}$$
$$\beta_{0j} = \gamma_{00} + \mu_{0j} \tag{16}$$

The equation does not contain an independent variable. Its role is to understand whether the variance of the dependent variable is not zero when the first level and the second level do not join the independent variable. The variance of the first and second level random effects is calculated by the cross level correlation coefficient ρ, which determines that the proportion of the total variation in the dependent variable is due to the difference in the second level, and the calculation equation of the coefficient ρ is that:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{17}$$

The cross-correlation coefficient ρ is between 0 to 1, and its numerical value reflects the proportion of the second level variation in the total variation of the dependent variable. (2) The random effect regression equation is established, and the independent variable is introduced at the first level of the random effect regression equation, while the second level of the equation has no independent variable. The function of the equation is to find the variation of the first horizontal intercept and the slope at the second level, so the random effect regression equation mainly focuses on the variance τ_{00} and τ_{1i} of μ_{0j} and μ_{1j}, to determine whether there is variation at the second level of the intercept and slope at the first level. The regression equation of the random effect is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} \times X_{1ij} + R_{ij}$$
$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10} + \mu_{1j} \tag{18}$$

(3) To establish a complete model based on the unconditional equations and random effects regression equations. The role of the complete model is to understand how the total variation of the dependent variable is affected by the first and second independent variables. When there is only one independent variable at the first level and the second level, the model is the simplest complete model, which is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + R_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times W_{1j} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times W_{1j} + \mu_{1j} \tag{19}$$

In the equation, γ_{00} and γ_{10} are fixed effects of β_{0j} and β_{1j}, which remain fixed in the second level units. γ_{01} and γ_{11} are regression coefficients of the second levels respectively, indicating the average change of β_{0j} and β_{1j} when the independent variable W_{1j} is changed per unit. μ_{0j} and μ_{1j} are random effects, the second level units are in the changing state, and their variance is τ_{00} and τ_{11}. τ_{00} is the second level independent variable to explain the residual variance after β_{0j}, τ_{11} is to explain the residual variance of β_{1j} with the second level independent variable. The covariance matrix of the second horizontal random effects is calculated as follows:

$$Var(\mu_{0j} \quad \mu_{1j}) = \begin{matrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{matrix} \tag{20}$$

When the second levels have multiple independent variables, the complete model is more complicated. It is known by equation (19) that when the first independent variable of the second level, W_{2j}, W_{3j} ..., are the second, third... independent variables of second level in turn. In addition, the independent variables contained in the equations of the second level can be the same or different.

A double-level model with two categorical variables as the dependent variable is constructed. The first level equation is similar to the Iogistic regression model when the dependent variable of the first level of the model is two categorical variables.

The first step is to determine the basic form of the model. The calculation equation of the first level is designed to be:

$$P(y_{ij}=1)=\pi_{ij}$$

$$Y_{ij}=\ln[\pi_{ij}\div(1-\pi_{ij})]$$

$$Y_{ij}=\beta_{0j}+\beta_{1j}X_{1ij} \tag{21}$$

The calculation equation of the second level is designed to be:

$$\beta_{0j}=\gamma_{00}+\mu_{0j}$$

$$\beta_{1j}=\gamma_{10}+\mu_{1j} \tag{22}$$

If the English examination for the first I student in the j university is passed (or excellent), then $y_{ij}=1$ and the student failed to pass the English examination, then $y_{ij}=0$, π_{ij} is the probability of passing (or excellent) for the English examination of the i student in university j. Y_{ij} represents the function of the dependent variable of the third student of university j.

The second step is to establish a double-level model with two categorical variables as dependent variables. The modeling process is basically the same as the multilevel model with dependent variables as numerical variables but the parameter estimation methods are different. There are two kinds of parameter estimation methods: The restricted Quasi Likelihood Estimation and the marginal Quasi Likelihood method. The former is first to estimate the regression coefficient of the first, and second levels by variance and covariance, and then make inferences based on the combined posterior groups. The latter is based on the Laplace transformation, and uses the likelihood function to estimate the parameters. The model established is:

$$Y_{ij}=\beta_{0j}+\beta_{1j}X_{1ij}+\pi_{ij}$$

$$\beta_{0j}=\gamma_{00}+\mu_{0j}$$

$$\beta_{1j}=\gamma_{10}+\mu_{0j} \tag{23}$$

3.2 The Construction of the Multilevel Factor Analysis Module

Based on the 2 double-level model established in 2.1, the two models are integrated to establish a multilevel analysis model for the performance factor of the English examination.

$$Y_{ij}=\beta_{0j}+\beta_{1j}X_{1ij}+R_{ij}+\pi_{ij}$$

$$\beta_{0j}=\gamma_{00}+\gamma_{01}W_{1j}+\mu_{0j}$$

$$\beta_{1j}=\gamma_{10}+\gamma_{11}W_{1j}+\mu_{0j}+\mu_{1j} \tag{24}$$

To sum it up, the establishment of a factor analysis model based on the multilevel environment for examination results is achieved.

4 EXPERIMENTAL CASES

4.1 Experimental Parameters

IN this study, there are 20 universities and 6265 university students in a total of 2003 full-time undergraduates, of which 4408 have participated in the CET-4 Examination with an examination rate of 75.4%. The highest score, the lowest score, the average number and the standard deviation of the first achievement were 99, 25, 65.84 and 13.10 respectively. The passing rate, the excellent rate and the passing rate were 69.86%, 5.045% and 52.7% respectively. The 20 university students' CET-4 scores and their passing rate, excellent rate, examination taking rate and passing rate are shown in Table 1. This study was stratified by the stratified cluster sampling and stratified by the cluster sampling. The information about each student's sex, age, nationality, Institute, subject type, time and achievement of the first time to take part in CET-4 level was collected. The results of their passing rate, excellent rate, examination rate and passing rate are the dependent variables, and the independent variables are gender, age, nationality, school system, examination term and subject type. The experimental data collected is shown in Tables 1 to 7.

Table 1. **The First CET-4 Examination Scores and Related Rates of 20 Universities (%).**

Number	Number of People	Scores	Passing rate	Excellence Rate	Examination Rate	Passing Rate
1	230	68.60±13.40	71.7	11.7	91.6	65.7
2	76	60.62±11.57	56.6	0.0	48.8	27.4
3	224	65.05±12.16	69.2	2.7	95.3	66.0
4	280	72.82±12.16	87.1	14.3	97.2	84.7
5	103	69.82±11.13	86.4	8.7	95.4	82.4
6	431	69.30±12.11	79.4	8.6	98.6	75.7
7	278	68.29±12.95	74.1	11.2	67.9	73.0
8	199	64.08±11.36	65.8	2.0	94.1	44.7
9	270	63.32±13.03	65.6	3.7	97.7	61.7
10	336	69.08±12.49	74.4	9.8	83.0	72.7
11	191	60.45±11.06	52.9	0.5	67.2	43.9
12	197	64.27±11.38	72.1	2.0	98.4	48.5
13	244	65.02±11.46	68.0	3.7	95.9	66.9
14	118	65.78±12.16	70.3	1.7	71.4	67.5
15	140	58.40±11.94	52.5	0.0	58.1	37.5
16	176	66.49±10.88	73.3	2.8	60.7	42.6
17	202	62.95±12.38	60.8	4.9	62.4	36.9
18	108	63.95±12.19	64.8	1.9	76.6	40.5
19	141	67.83±13.21	75.9	5.7	33.5	58.2
20	464	65.49±13.15	76.3	5.0	72.9	25.5
amount to	4408	65.84±13.10	69.86	5.045	78.335	56.1

It can be seen from Table 1 that among the 20 colleges, the College of Computer Science and Technology (No. 6) has the highest examination rate, and the Medical School (No. 19) has the lowest examination rate. The average score, the passing rate, the excellent rate and the passing rate of the students of the College of Life Sciences (No. 4) are the first in 20 colleges and the average score and the passing rate are all. The colleges with the lowest rate of passage are the School of Physical Education (No. 20), the School of Art (No. 2) and the Medical School (number 19).The 20 colleges have a low rate of excellence. The highest college is 14.3%, and the excellent rate of 2 colleges is 0. By comparison, the difference between the scores of the first English examination scores of the 20 colleges was $F=28.273$, $P<0.001$, and the passing rate, the excellent rate, the examination rate and the passing rate were 409.505, 145.617, 2344.333 and 1335.505 respectively, and their P values were all less than 0.001.

This study describes the distribution characteristics of the first grade English examination scores by the means of average, standard deviation, median, mass, passing rate, excellent rate, examination rate. The scores of 60 points and above are passed, 85 and above are excellent. The calculation equations of the passing rate, the rate of excellence, the examination rate are as follows:

The Passing rate = the number of students whose examination scores exceed 60 points ÷ the total number of candidates taking part in the examination × 100%.

The Excellent rate = the number of students whose examination scores exceeded 85 points ÷ the total number of candidates taking part in the examination × 100%.

The Examination rate = the number of persons participating in the examination ÷ the total number × 100%.

The Passing rate = the number of students whose examination scores exceeded 60 points ÷ the total number = 60 or more × 100%.

Table 2. **The Correlation between Gender and First Grade Examination Scores in the CET-4 Examination (%).**

Sex	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing rate
Male	64.49±12.99	65.6	4.7	80.6	52.9
Female	69.02±12.79	80.2	9.2	65.4	52.4

As can be seen from Table 2, the examination rate of girls is lower than that of boys but their scores, passing rates and excellent rates are higher than those of boys. P is less than 0.05.

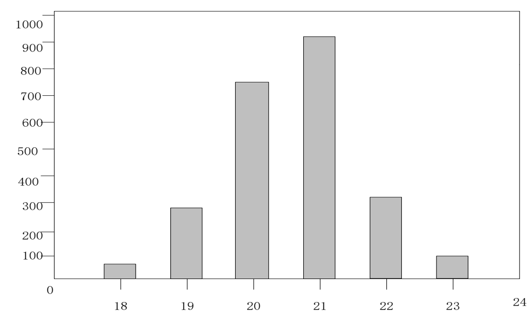

Figure 1. **The Age Distribution Chart for Students who Participated in the CET-4 Examination.**

Among the 4408 students, the youngest was 18 years old, and the oldest was 23 years old. The average age was 20.6 years, and the median and mode were 21 years. It is shown in Table 3 that students younger than 20 years old are better than the 20-year olds, and $P<0.05$. This data indicates that the English proficiency of the pupils is higher than that of the older pupils.

Table 3. **The Correlation between Age and First Grade Scores in the CET-4 examination (%).**

Age	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing rate
≤20	69.80±12.34	79.4	10.2	74.2	58.9
>20	61.90±12.63	60.6	1.9	76.6	46.4

From Table 4, we can see that scores, passing rate of the CET-4 examination for the Han students are higher than that of the minority students, and $P<0.05$.

Table 4. **The Correlation between Nationality and the First Examination Scores of the CET-4 (%).**

Nationality	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing Rate
Han nationality	67.88±8.45	93.8	8.4	41.5	39.0
Minority nationality	65.65±9.17	93.9	6.1	22.8	21.5

It can be seen from Table 5 that the four-year students have the highest scores of the passing rate, excellent rate, examination rate. The six-year students have the lowest passing rate, excellent rate, examination rate except for the examination scores. By comparison, there was no difference of the five indexes between the six-year students and the seven-year students, P>0.05. Except for the excellent rate, the other indexes of the six or seven-year students were all lower than the five-year students, P<0.01, and the other indexes of the five-year students were all lower than the four-year students, P<0.01 except for the passing rate. The data shows that in the CET-4 examination, there are some differences in the five indicators among the students who are not in the same school system.

Table 5. **The Correlation Rate between the School System and First Examination Scores of the CET-4 (%).**

Educational system	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing Rate
4	66.04±13.15	70.5	6.3	82.7	58.3
5	63.01±10.90	64.4	1.6	52.7	33.9
6	55.50±10.52	12.5	0	4.7	0.6
7	55.07±10.31	32.1	0	9.4	3.0

The examination rate and the passing rate of the students in the fourth term are the highest. The scores, passing rates and excellent rates in the second semester are higher than those of the other semesters. The average passing rate of the examination scores gradually decreases with the increase of the term, and the students' examination results and reference time are gradually increased. The negative correlation, r_s =-0.490, and P<0.01, that is, students who attend the English examination earlier, their English level is higher and the difference between students is smaller. On the other hand, students who attended the English examination later are poor in English and the difference between students is larger.

Table 6. **The Correlation Rate between the Examination Term and the First Examination Scores of the CET-4 (%).**

Examination Term	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing Rate
2	80.29±9.81	94.6	32.6	3.9	3.7
3	77.29±9.31	94.6	21.0	12.7	12.0
4	62.84±11.55	63.8	1.1	66.9	42.7
5	58.87±15.56	56.9	2.9	7.9	4.5
6	53.29±18.25	44.1	3.4	3.7	1.6

It is seen from Table 7 that the class of science and engineering is higher, literature history and art are in the middle, and the medical class is the lowest. From the other four indexes, the students of science, engineering, literary history and medicine are higher than the stylistic arts. It seems that the students of art have the lowest English level in 20 colleges.

Table 7. **The Correlation between the Subject Types and the First Examination Scores of the CET-4 (%).**

Subject Type	Scores	Passing Rate	Excellence Rate	Examination Rate	Passing Rate
science	68.50±12.93	76.2	9.0	87.9	67.0
engineer	66.14±12.43	70.6	5.8	88.7	62.6
liberal	66.16±12.57	71.6	5.6	694	49.7
medical	63.02±11.88	64.0	2.9	30.3	19.4
art	42.85±10.81	6.6	0	70.5	4.6

4.2 Experimental Results

In this study, the standard error of the calculation results is used as the standard. The traditional algorithm and the multilevel algorithm are used respectively. The results are compared with the experimental results. The experimental results are shown in Table 8 and 9.

Table 8. **The Calculation Results of the Traditional Algorithm.**

Variable		Parameter Estimation	Standard Error	T	P
Nodal Increment		57.636	1.1673	49.38	<0.001
Sex	Male	-4.953	0.3628	-13.65	<0.001
	Female				
Age	>20	-3.006	0.337	-8.91	<0.001
	≤20				
Nationality	Han	4.478	0.773	5.79	<0.001
	Minority				
Educational System	4				
	5	-2.382	1.012	-2.30	0.021
	6	-6.729	3.902	-1.72	0.045
	7	-7.960	2.240	-3.55	<0.001
Examination Term		-7.568	0.273	-27.72	<0.001
Subject type	Science	20.426	1.058	19.31	<0.001
	Engineer	20.949	1.010	20.74	<0.001
	Liberal	21.200	1.025	20.69	<0.001
	Medical	20.318	1.372	14.81	<0.001
	Art				

Table 9. **The Calculation Results of the Multilevel Algorithm.**

Variable		Parameter Estimation	Standard Error	T	P
Nodal Increment		63.342	0.681	93.033	<0.001
Sex	Male	-4.628	0.364	-12.711	<0.001
	Female				
Age	>20	-2.499	0.332	-7.533	<0.001
	≤20				
Nationality	Han	4.306	0.758	5.681	<0.001
	Minority				
Educational System	4				
	5	-1.774	1.348	-1.316	0.048
	6	-0.261	5.555	-0.047	0.036
	7	-4.016	3.330	-1.206	0.022
Examination Term		2.848	0.329	8.662	0.000
Subject Type	Science	24.607	3.106	7.922	<0.001
	Engineer	22.618	2.909	7.973	<0.001
	Liberal	21.583	2.953	7.310	<0.001
	Medical	18.909	3.165	5.974	<0.001
	Art				

According to experimental results in Tables 8 and 9, we can draw the standard error comparison diagram of the traditional algorithm and the multilevel algorithm, and Figure 2 is shown in detail.

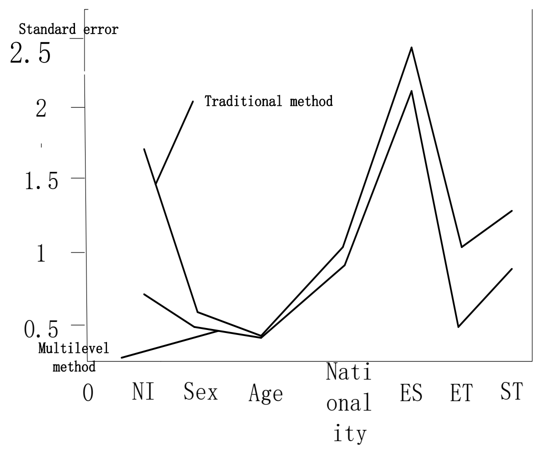

Figure 2. The Standard Error Comparison of the Experimental Results between the Traditional Algorithms and the Algorithm in this Study.

From Figure 2 and Tables 8 and 9, we can see that the algorithm proposed in this study has obvious advantages over the traditional algorithm in the error of the calculation results. The standard error of the calculation results in this study is lower than that of the traditional algorithm, which shows that the results of this algorithm are more accurate in the process of the analysis. It is very effective to analyze the influential factors of University English.

5 ANALYSIS OF THE EXPERIMENTAL RESULTS

ACCORDING to the experimental results of 3.2, we can see that the P values of gender, age, nationality, school system, examination term and subject type are less than 0.05, that is, the six factors have great influence on the University English performance, so this study analyzed the six factors in turn.

5.1 The Influence of Gender on the English Proficiency Examination

Despite the overall level of intelligence, there is very little gender difference in the intelligence development between boys and girls. But there are some differences between boys and girls in their way of thinking. Normally, boys are good at abstract thinking, rational thinking, common sense, similarity and mapping. Girls are good at image thinking, perceptual thinking, retelling numbers and decoding. English is a more memorable subject, and girls have more advantages. In addition, the girls' desire for learning knowledge is obviously stronger than that of boys, with clear learning goals and hard work, and the main energy is put into the study. Therefore, under normal circumstances, female students in universities are better than boys. This study analyzed the relationship between the scores of the CET-4 examination and gender, and the results showed that "gender" is an important factor affecting the achievement of the English proficiency examination. Therefore, the students of different sexes should adopt different teaching methods to correctly guide them to set up a good study style and cultivate their interest in English learning, which will be more conducive to the improvement of the students' English level.

5.2 The Influence of the Age on the English Proficiency Examination

This study shows that students who are below 20 years old have higher scores, passing rates and excellent rates than those over 20 years of age in the CET-4Examination. This is due mainly to the fact that most of the younger students are in the first or second grade of the University. At this time, there is a special explanation of the English teachers, and their listening, speaking, reading and writing are greatly improved. They also want to pass the examination, so the younger students have a better examination. The older students lack the learning of the English system, and the later they take the examination, the less confidence they have and the worse the examination results are. Therefore, in English teaching, we must help students establish confidence in the examinations, seize the characteristics of students at different ages, and increase their teaching efforts.

5.3 The Influence of the Nationality on the English Proficiency Examination

Ethnic minorities have their unique national psychological characteristics due to their unique historical and cultural origins and their living environment. Minority students have good personality traits such as hard work and self-discipline, but they also have bad personality psychology such as depression, anxiety and interpersonal sensitivity. In addition, they are limited by the English teaching level in minority areas. Their English entrance level is low, and the University English teaching does not aim at minority students. The specific situation of targeted teaching will affect their examination results. This study shows that there is a great difference between the scores and the passing rates of the English proficiency examination of the Han and ethnic minority students, especially the passing rate. Therefore, on the basis of understanding and grasping the personality of the minority students, we should strengthen their education and guidance, help them overcome the interference of bad moods and so on, in order to improve their mental health quality, and take this as a breakthrough to improve their cultural quality.

5.4 The Influence of the Academic System on the English Proficiency Examination

Four-year students have a short time in school. They think that time is relatively urgent, and the learning pressure is relatively high. Students who have longer academic years have more opportunities to take part in the English proficiency examination consciously, so their English learning is more relaxed, and their examination results are relatively low. This study shows that the four-year students who take the English proficiency examination have the highest number of students. Therefore, in English teaching, we should seize the characteristics of the students, effectively supervise their study, and encourage them to pass the examination in an air of breath in order to lighten the burden of later studies.

5.5 The Influence of the Examination Term on the English Proficiency Examination

In general, only the students with high English levels can take the English proficiency examination in the second term. At the present time, they have more English class hours and a study system, while other students usually take the examination in the fourth semester, so the semester in the English proficiency examination reflects the students' English level to a certain extent. This study shows that the examination term is an important factor affecting the performance of the English proficiency examination. There is a negative correlation between the students' examination results and the time of reference. That is, students who attend the English examination earlier are higher in English and smaller among students. On the other hand, students who attend the English examination later are poor in English and the difference between students is larger. Therefore, universities should make targeted University English teaching plans according to their own characteristics, teach students in accordance with their aptitude, and effectively improve their English level.

5.6 The influence of subject types on the English proficiency examination

This study shows that the subject types are the factors affecting the performance of the English proficiency examination. Compared with the science, technology, history and medicine, students of stylistic and artistic have a lower level of English. Therefore, in English teaching, we should seize the characteristics of the different types of students, strengthen the English teaching of stylistic and artistic students, tamp their English foundation and improve their English level?

6 CONCLUSION

THE education measurement and teaching evaluation is an important part of teaching activities. It is an important means to carry out scientific management in the field of teaching and an effective guarantee to improve the quality of education. One of the main contents of the teaching evaluation is the analysis of the examination results and its influencing factors. At present, descriptive analysis, cluster analysis, principal component analysis, factor analysis and multiple linear regression analysis are the main methods used to analyze the results of the examination and its influencing factors. However, in the teaching evaluation, it is often necessary to analyze hierarchical structure data, such as examination results of different schools or students of different majors in the same school. Data with hierarchical structures are not suitable for analysis by the above methods. In this study, based on the determination of data, estimation methods and latent variables, the environment of the multi-level analysis model is established. By enumerating examples, the effectiveness of the multi-level model in the analysis of the influence factors of the University English performance is verified. It is hopeful that this article can provide help for related research.

7 ACKNOWLEDGMENT

THIS paper is supported by the special fund of "young and middle-aged teachers development program for visiting scholars from Jiangxi Undergraduate colleges and universities (No. 109, [2016], Jiangxi Education Department Office Document)", which is one of the achievements.

8 REFERENCES

Chen Jinyang. Research on Enterprise responsibility rating based on Multi-level and Multi-class logistic Model[D]. Yunnan University of Finance and Economics.2017,15(02):111-113.

Cheng Yong. Application of variable selection of multilevel model in per capita income data of farmers [D]. Yunnan University of Finance and Economics.2015,32 (12): 102-106.

Cui Junying. Research on health trend and influencing factors of bingtuan resident’s self-assessment based on multi-level model [D]. Shihezi University.2015,31(12): 37-40+97.

Dai Jia is a man. A multi-level analysis of oral health factors of 12-year-old children in Chongqing based on Bayes. Chongqing Medical University, 2017,15(22):110-111.

Ding Cheng. Application of different statistical models in public health research [D].Zhejiang University. 2015,31(06): 443-446.

Guo Dong. Study on the selection of key influencing factors on hospitalization cost of single disease [D]. Kunming University of Technology.2017,33(06):193-196.

HaoGuang. Association Study of Sleep and Cardiovascular Diseases, Diabetes Mellitus, Beijing Institute of Peking Union Medical University, 2015 ,31(11): 245-248+260.

Hu Jianxiang. Study on the relationship between Dietary Fat intake and Obesity related Indexes and Blood pressure [D]. China Centers for

Disease Control and Prevention (CDC).2017,33(08):205-209.
Hu Liang. Research on Modeling method of recommendation system integrating multiple Information [D]. Shanghai Jiaotong University.2015,31(12): 142-146 / 173.
Huang Shafei. Study on hospitalization cost and influencing factors of a third-Class A Hospital based on Recursive system Model [D]. Guangxi Medical University.2017,33(07):305-308+3.
LengQingfan, Bo Yuming, Zhao Gaopeng, Liu Lujiang. Research on Precision estimation of Rendezvous and docking position and attitude of Space spacecraft [J]. Computer simulation: 2016,33(12): 47-53.
Liu Jie Zhen. A study on the influence of Social Security on consumption in Western Rural areas [D]. Yunnan University of Finance and Economics.2017,33(02):147-152.
Liu Miao. Study on prevalence and incidence of type 2 diabetes mellitus and risk model of type 2 diabetes mellitus in middle aged and old people in 5 years [D]. Chinese people's Liberation Army Medical University, 2017,33(03): 32-36.
Lu Chenlong. Application of structural equation Model in Cognitive Aging and its influencing factors [D]. The Academy of military Medical Sciences of the Chinese people's Liberation Army (PLA), 2015,31(06): 168-172 200.
Qin Yin-yi. Research and Application of tool variable method based on Multi-level Model [D]. The second military Medical University.2015,30(02): 247-251.
Sun Lin. Complex sampling and Multi-level Model-Health Expenditure per capita and two-week prevalence in Shaanxi Province [D]. Fourth military Medical University, 2015,31(01): 253-257.
Sun Xiaojian. Multi-level side model of grade response and its application in subjective problem scoring [D]. Zhejiang normal University.2016,34 (04): 302-305+322.
Wang Lanlan. Effects of family income and educational level on BMI and its changing characteristics in middle-aged residents [D]. Zhengzhou University.2017,33(03):288-291+309.
Yu Feifei. Construction and Application of Multi-level tendency scoring Model for classified data [D].Second military Medical University. 2016,31(10): 241-245.
Yu Xuejiao. Study on the factors of Energy intensity difference in China [D]. Yunnan University of Finance and Economics, 2015,34 (08): 351-355.

9 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

10 NOTES ON CONTRIBUTORS

Shaoyun Long, Ji'an, Jiangxi, Master, associate professor, Corpus linguistics; corpus translation study, Foreign Languages College, Jiangxi Normal University, 330022.

Qianying Long, Ji'an, Jiangxi, Ph.D., Tourism service and data translation, College of Tourism and Service Management, Nan-kai University, 300350.