ISSN : 0974 - 7435 Volume 10 Issue 11

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(11), 2014 [5433-5441]

Research and analysis of prediction model of English level of modern college students

Shuying Han

School of Foreign Languages, Handan College, Handan 056005, Hebei, (CHINA)

E-mail : shuyinghan_2014@yeah.net

ABSTRACT

With increasingly frequent cooperation between all countries all over the world, as one of the most common languages, English has a more and more important position. The continuous improvement of productivity intensifies the competition of all kinds of industries in the society, so most of the enterprises continuously improve the recruitment threshold upon employee recruitment, in which one of the conditions is that the employee must pass CET4 and CET6. Based on this, in this paper, the English level of college students is studied according to the CET4 and CET6 scores of investigated college students. Then, the English level of college students is analyzed by establishing the curve fitting model. Finally, plotting and prediction are conducted according to the existing data by establishing the fitting function with the least squares method and using MATLAB for quadratic fitting. It is concluded through analysis that the English level of college students are gradually increased with the increase of the passing rate of CET4 and CET6 of college students.

KEYWORDS

CET4; CET6; Curve fitting; Least squares method; MATLAB.

INTRODUCTION

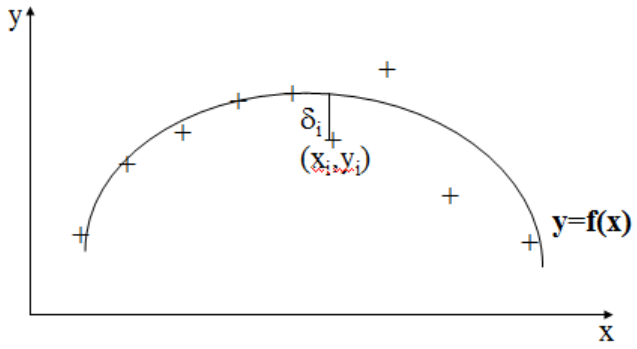
With increasingly frequent cooperation between all countries all over the world, as one of the most common languages, English has a more and more important position. The continuous improvement of productivity intensifies the competition of all kinds of industries in the society, so most of the enterprises continuously improve the recruitment threshold upon employee recruitment, in which one of the conditions is that the employee must pass CET4 and CET6. In addition, due to the increased college graduates in China for recent years, college students face a tough employment situation and unprecedented challenges. Therefore, to increase the employment opportunity, many college students take every chance to get certificates, especially CET4 and CET6 Certificate. Considering that CET4 and CET6 certificates are more and more popular, whether CET4 and CET6 scores can represent the English level of college students becomes a hot topic.

Based on this, in this paper, the English level of college students is studied according to the CET4 and CET6 scores of investigated college students. Then, the English level of college students is analyzed by establishing the curve fitting model. Many scholars have conducted researches on the curve fitting, for example: an improvement of least squares method in the complex field by Li Mengxia, Chen Zhong; algorithm implementation of classification problems of least squares support vector machine by Zhou Jianping, Zheng Yingping and Wang Zhiping. Finally, plotting and prediction are conducted according to the existing data by establishing the fitting function with the least squares method and using MATLAB for quadratic fitting.

Based on the application of curve fitting model by precedents, CET4 and CET6 scores are analyzed to get the English level of college students, thus making contributions to the promotion of CET and CET6.

LEAST SQUARES METHOD

On the two-dimensional plane, n points $(x_i y_i),(i=1,2,...n)$ are known and $y=\varphi(x)$ is found, in which $\varphi(x)$ approaches the points in the figure under certain condition, that is, the curve fitting is good.



Where, δi is the distance between point (*x*i,*y*i) and curve *y*=*f*(*x*).

$\delta_i = y_i - y_i^*$ is the residual, in which y_i^* is the approximate value and y_i is the observed value.

The smaller the residual is, the better $\varphi(x)$ will be; the larger the residual is, the worse the $\varphi(x)$ will be. The three recognized criteria are as follows:

(1) Make the sum of the absolute value of residual minimal, that is, $\sum_i |\delta_i| = \min$

(2) Make the maximum absolute value of residual minimal, that is, $\max_i |\delta_i| = \min$

(3) Make the quadratic sum of residual minimal, that is, $\sum_i \delta_i^2 = \min$

The method of calculating $\varphi(x)$ by criterion (2) is called best uniform approximation of function. Due to the complexity of the calculation and application of absolute value, criterion (1) is rarely used. The method of calculating $\varphi(x)$ by criterion (3) is called best square approximation of function, that is, the least squares method of curve fitting, so this criterion is called least squares criterion.

The least squares method is defined as: given the data set (x_i, y_i) $(i = 1,2, \cdots, n)$, it is found the approximate function $\varphi(x)$ satisfies the condition that the following equation is minimum:

$$\sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [y_i - \varphi(x_i)]^2$$

This method of calculating $\varphi(x)$ is called least squares method of curve fitting and $\varphi(x)$ is called the corresponding least squares function.

A set of data (x_i, y_i) $(i = 1,2,\cdots, n)$ are known, a m $(m < n)$ polynomial is obtained

$$P_m(x) = \sum_{j=0}^{m} a_j x^j$$

To make the quadratic sum of error minimal

$$\sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [y_i - P_m(x_i)]^2$$

$P_m(x)$ is the least square m fitting polynomial of this set of data.

First, a set of functions $r1(x)$, $r2(x)$, …$rm(x)$, m<n, are selected,

$$f(x) = a_1 r_1(x) + a_2 r_2(x) + ... + a_m r_m(x)$$

Where, $a1$,$a2$, …am are undetermined coefficients.

Criterion of calculating $a_1, a_2, \cdots, a_m$ (least squares criterion): for n points (x_i, y_i) on the two-dimensional plane, make the quadratic sum of δ_i minimal, in which δ_i is the distance between n points (x_i, y_i) and curve $y = f(x)$.

$$J(a_1, a_2, \cdots a_m) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [f(x_i) - y_i]^2 = \sum_{i=1}^{n} [\sum_{k=1}^{m} a_k r_k(x_i) - y_i]^2 ,$$

Then, calculate $a_1, a_2, \cdots, a_m$ and $J(a_1, a_2, \cdots, a_m)$ is minimal.

$$\frac{\partial J}{\partial a_k} = 0 \quad (k = 1, \cdots m) \Rightarrow \begin{cases} \sum_{i=1}^{n} r_1(x_i)[\sum_{k=1}^{m} a_k r_k(x_i) - y_i] = 0 \\ \cdots\cdots \\ \sum_{i=1}^{n} r_m(x_i)[\sum_{k=1}^{m} a_k r_k(x_i) - y_i] = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum_{k=1}^{m} a_k \sum_{i=1}^{n} r_1(x_i) r_k(x_i) = \sum_{i=1}^{n} y_i r_1(x_i) \\ \cdots\cdots \\ \sum_{k=1}^{m} a_k \sum_{i=1}^{n} r_m(x_i) r_k(x_i) = \sum_{i=1}^{n} y_i r_m(x_i) \end{cases}$$

The normal equation corresponding to polynomial fitting is:

$$\sum_{k=0}^{m} a_k \left(\sum_{i=0}^{n} x_i^{k+j} \right) = \sum_{i=0}^{n} y_i x_i^j \quad (j = 0,1,...,m)$$

$$A^T AX = A^T Y$$

The above equation can be obtained.

Where, $\mathrm{A} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^m \end{bmatrix}$, $X = \begin{bmatrix} a_0 \\ a_1 \\ \cdots \\ a_m \end{bmatrix}$.

For example, the linear fitting of discrete data is conducted according to the least squares method as follows:

$\mathbf{x_i}$	**-1.00**	**-0.50**	**0**	**0.75**	**1.00**
y_i	0.2200	0.8000	2.0000	2.5000	3.7500

Suppose that the fitting polynomial is $p(x) = a_0 + a_1 x$, equation: $AX = Y$, so

$$\begin{bmatrix} 1 & -1.00 \\ 1 & -0.50 \\ 1 & 0 \\ 1 & 0.75 \\ 1 & 1.00 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 0.2200 \\ 0.8000 \\ 2.0000 \\ 2.5000 \\ 3.7500 \end{bmatrix}$$

Normal equation $A^T AX = A^T Y$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1.00 & -0.50 & 0 & 0.75 & 1.00 \end{bmatrix} \begin{bmatrix} 1 & -1.00 \\ 1 & -0.5 \\ 1 & 0 \\ 1 & 0.75 \\ 1 & 1.00 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1.00 & -0.50 & 0 & 0.75 & 1.00 \end{bmatrix} \begin{bmatrix} 0.2200 \\ 0.8000 \\ 2.0000 \\ 2.5000 \\ 3.7500 \end{bmatrix}$$

After simplification,

$$\begin{bmatrix} 5 & 0.25 \\ 0.25 & 2.8125 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 9.45 \\ 5.005 \end{bmatrix}$$

It is calculated that $a_0 = 1.80906$, $a_1 = 1.61875$, so the fitting polynomial is as follows:

$$p(x) = 1.80906 + 1.61875x .$$

ANALYSIS OF WRITING LEVEL

A model is established according to CET4 scores of college students recently:

The statistics (with the average scores of four provinces as the source data) are shown in TABLE 1:

TABLE 1 : Statistics of CET4 writing scores of college students in four provinces

Province	1990	1995	2000	2005	2010
A	123.6	99.8	104	112	89.4
B	90	117.2	110.8	99.6	109.6
C	105.4	95.4	114.8	108	101
D	96	120.6	95.4	95.8	112.2

Then, according to the data in the above table, the following figure is obtained by fitting with MATLAB (see Figure 1) (The blue dotted line is the first fitting curve, the green dot dash line is the second fitting curve and the red solid line is the third fitting curve).

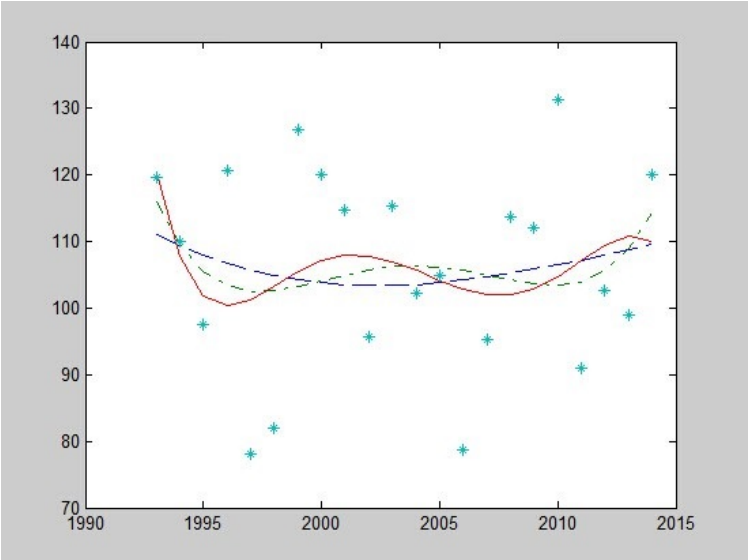


Figure 1 : Fitting curve of the writing level of college students

According to the curve in Figure 1, the writing level of college students at present is in a downward trend.

ANALYSIS OF READING LEVEL

The statistics (with the average scores of four provinces as the source data) are shown in TABLE 2:

TABLE 2 : Statistics of CET4 reading scores of college students in four provinces

Province	1990	1995	2000	2005	2010
A	127.4	125	105.4	129.2	100.2
B	131.2	141.4	144.8	120	123.4
C	122.8	139.2	136.8	136	137
D	107.2	139	116.2	110	140.8

Then, according to the data in the above table, the following figure is obtained by fitting with MATLAB (see Figure 1) (The blue dotted line is the first fitting curve, the green dot dash line is the second fitting curve and the red solid line is the third fitting curve).

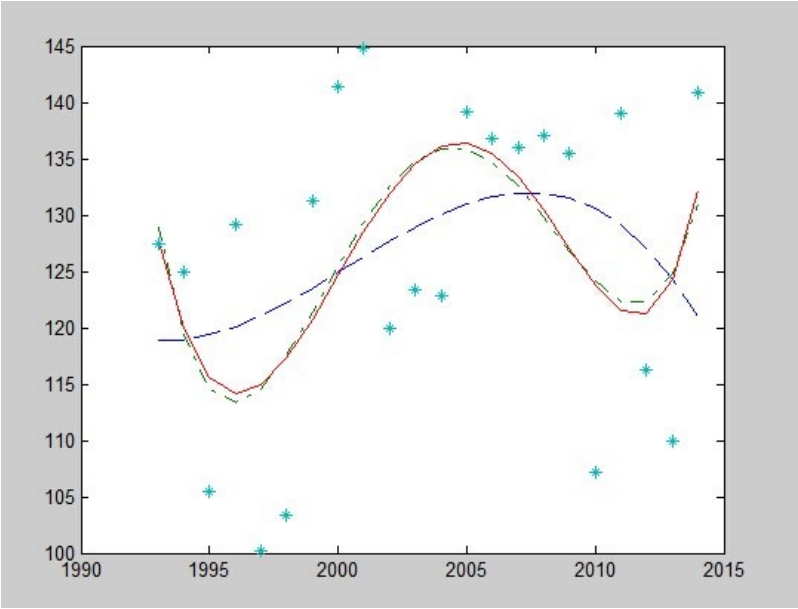


Figure 2 : Fitting curve of the reading level of college students

According to the curve in Figure 2, the reading level of college students at present is in an upward trend.

ANALYSIS OF LISTENING LEVEL

The statistics (with the average scores of four provinces as the source data) are shown in TABLE 3:

TABLE 3 : Statistics of CET4 listening scores of college students in four provinces

Provinces	1990	1995	2000	2005	2010
A	99	102.6	91	131.4	112
B	95.2	78.8	104.8	102.2	115.4
C	95.6	110	119.6	120.6	97.6
D	113.4	110	119.6	124.	115.6

Then, according to the data in the above table, the following figure is obtained by fitting with MATLAB (see Figure 3) (The blue dotted line is the first fitting curve, the green dot dash line is the second fitting curve and the red solid line is the third fitting curve).

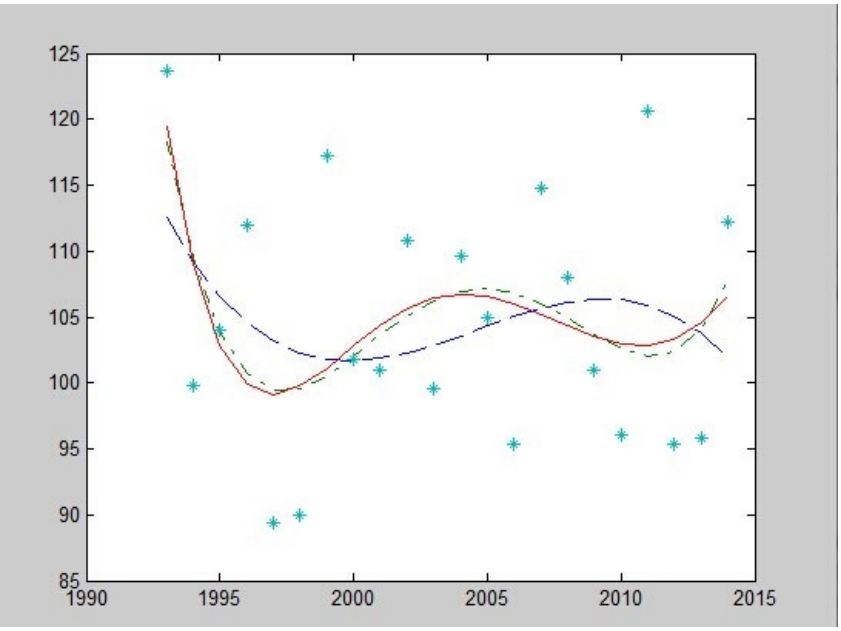


Figure 3 : Fitting curve of the listening level of college students

According to the curve in Figure 3, the listening level of college students at present is in a downward trend.

ANALYSIS OF OVERALL LEVEL

Then, according to the data in the above table, the following figure is obtained by fitting with MATLAB (see Figure 4) (The blue dotted line is the first fitting curve, the green dot dash line is the second fitting curve and the red solid line is the third fitting curve).

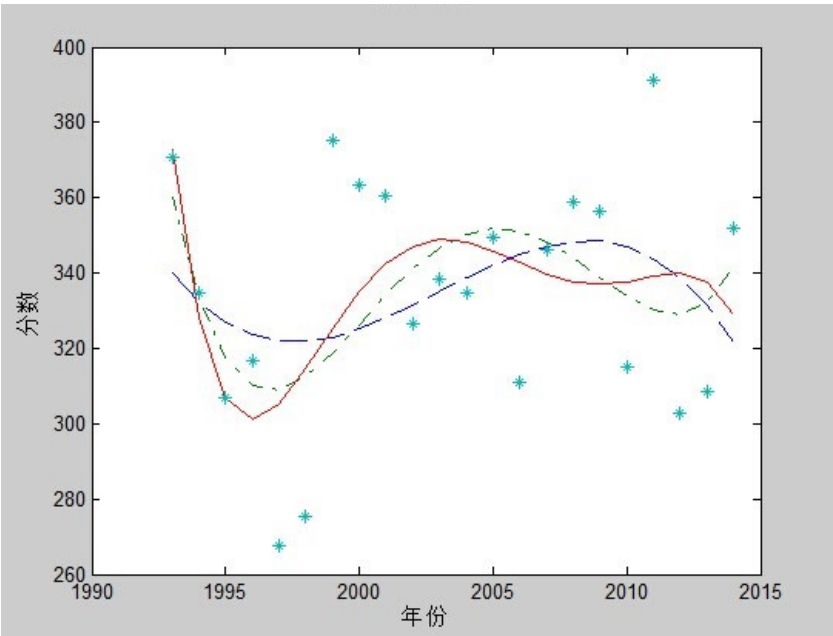


Figure 4 : Fitting curve of the overall level of college students

According to the curve in Figure 4, the overall level of college students at present is in a downward trend.

CONCLUSION

In this paper, according to the single and overall scores of CET4 in four provinces from 1990-2015, the curve fitting model is established and then the solving is conducted by least squares method to make the conclusion that the English level of college students gradually declines. However, considering the high-score striver, scoring system and other objective factors of CET4 and CET6, the English level of college students cannot be evaluated only in accordance with CET4 and CET6 scores.

ACKNOWLEDGEMENTS

This project is supported by level-4 project for 2014 comprehensive reform of Handan College.

REFERENCES

[1] Wang Huiwen; Partial Least Squares Regression and Its Application[M]. National University of Defense Technology Press, **(1996)**.
[2] Liu Donghong; Deviation of Analytic Formula of Linear Function by Undetermined Coefficient Method[J]. Friends for Middle School Students, **35**, **(2008)**.
[3] Zuo Wanjuan, Yang Mengfei, Duan Yonghao; Computer Application and Software[M], **26(9)**, **(2009)**.
[4] Chen Yanguo; Analysis on Stability of Regression Forecast Model (Outlier Treatment)[J]. Engineering Geology Computer Application, **03**, **(2005)**.
[5] Gao Huisheng, Liu Tongna, Li Cong; Publishing House of Electronics Industry, **(2010)**.
[6] Liu Xiaoli, Chen Chunmei; Piecewise Curve Fitting Method Based on Least Squares Principle[J]. Journal of Yili Education College, **03**, **(2004)**.
[7] Jin Liming; Preliminary Discussion on Improving Business English Teaching Level[J]. Journal of Zhejiang Business Technology Institute, **2**, **(2003)**.
[8] Kang Hong; Discussion on Method of Improving Students' English Level[J]. Journal of Qinghai Nationalities Institute, **3**, **(2004)**.