East Asian Math. J.
Vol. 33 (2017), No. 4, pp. 467–482
http://dx.doi.org/10.7858/eamj.2017.032

ON COLLEGE MATHEMATICS EDUCATION BY COMPARATIVE ANALYSIS OF COLLEGE ACHIEVEMENTS AND SCORES OF COLLEGE ENTRANCE EXAMINATION

Guang-Ri Piao and Mei-Hong Cui

Abstract. This paper studies the relationship between the college entrance examination scores and college mathematics achievements of the freshman, who are the students of college of technology and management in Yanbian University, by employing statistical analysis. First of all, we start from the correlation analysis of the majors and regions of students. Secondly, with gender and region as the classification, we divide the sample into two parts and test the mean significance, and we find that not only different gender but also different region have significant differences to college mathematics achievements. Thirdly, we use the college entrance examination scores as independent variable, and college mathematics achievement as the dependent variable to establish regression model. The model provides a realistic basis for the education department of our school, and provides reference value for the other major research.

1. Introduction

In order to speed up the further development of education, and improve the quality of the people comprehensively, the government of China promulgated *National Education Plan* (in the brief, NEP). NEP pointed out that deepening educational reform and improving the system of evaluation of university teaching are very important.

The quality of teaching, which we have to mention here, is almost the key role in educational reform, and in addition, it is reflected by student achievement at school more or less. Therefore recently, the research and analysis of college students in school results have attracted much attention. In the process of analyzing the results of undergraduates in university, we have put forward the college entrance examination results and whether there is correlation among the results, and how big the correlation is, and whether the results can be accurately predicted by the college entrance examination students in school.

Received August 3, 2017; Revised August 22, 2017; Accepted August 28, 2017.
2010 *Mathematics Subject Classification.* 97C70; 97D50.
Key words and phrases. College entrance examination scores; College mathematics achievement; Statistical analysis; Teaching method.

As early as the last century 90's, Ferrari and Parker [5] predicted college student achievement by employing high school academic performance, self-efficacy and psychological control source. Busato *et al.* [1] did some forecast analysis on the academic situation of higher education in the aspects of knowledge ability, learning style, personality and achievement motivation. Duff *et al.* [4] studied the relationship between personality and studying by using linear regression analysis method.

In addition, there are also many domestic researchers focusing on the statistical analysis of college students in school performance, for example, see [2, 3, 6–9] and the references therein. Yu *et al.* [8] investigated the relationship between college entrance examination and college students' academic achievement by using *Pearson* linear correlation analysis and *Copula function* nonlinear correlation analysis, and the results showed that there is a weak correlation between them. Ding *et al.* [3] compared university achievement of college students in science and engineering of two 985 PROJECT universities, and the results showed that college entrance examination scores or single subjects and university performance was not a simple linear correlation, Only if the college entrance examination score is less than a certain threshold, university achievement was determined by the college entrance examination results and student characteristics.

In the present paper, we adopts a statistical analysis to reveal the relationship between the college entrance examination scores and college mathematics achievements of the freshman, who are the students of institute of technology and management in Yanbian University. We begin with the correlation analysis of the majors and regions of students. As a result, college entrance examination scores of engineering institute students are relative strongly related to college mathematics achievements. Then, with gender and region as the classification, we divide the sample into two parts and test the mean significance, and we find that not only different gender but also different region have significant differences to college mathematics achievements. After that, we use the college entrance examination scores as independent variable, and college mathematics achievement as the dependent variable to establish a regression model. The model provides a realistic basis for the education department of our university, and provides reference value for the other major research.

2. Preliminaries

In this section, we give some necessary preliminaries, which will be used in the later analysis.

2.1. Regression model and parameter estimation

(*Regression Model*) If variables $x_1, x_2, \ldots, x_p$ and random variable y are related, and assume independent variables $x_1, x_2, \ldots, x_p$ and dependent variable y are linear dependent, then the obtained $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ satisfy the following linear regression model:

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i, \quad (i = 1, 2, \ldots, n) \\ E(\epsilon_i) = 0, var(\epsilon_i) = \sigma^2, cov(\epsilon_i, \epsilon_j) = 0 (i \neq j) \\ \text{or } \epsilon_i \square N(0, \sigma^2), \text{ independent } (i = 1, 2, \ldots, n) \end{cases}$$

where y is called *explained variable* (or dependent variable), and $x_1, x_2, \ldots, x_p$ are called *explanatory variables* (or independent variables). In regression model, the regression function $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$ shows clearly the related relationship between variables $x_1, x_2, \ldots, x_p$ and random variable y, and the other part is random error ϵ. Actually from here we know that there is a certain relationship between regression analysis and correlation analysis, but they are different. Usually we in addition suppose that the n-tuple data are obtained independently, and so $y_1, y_2, \ldots, y_n$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are random variables independently each other.

(*Parameter Estimation*) If $rank(x) = k + 1 \leq n$, the least squares estimation (OLSE) of regression parameter β is

$$\hat{\beta} = (x'x)^{-1} x'y.$$

After we give the estimation $\hat{\beta}$ of β, and substitute into regression model and omit the error part, we obtain an equation, which is called unary regression equation, and denoted by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1, \tag{1}$$

then OLSE of the unary regression equation (1) is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{x}$ and $\bar{y}$ are the sample averages of x and y, respectively.

We call the difference of sample observation values y_i and regression fitting value $\hat{y}_i$ *residual*, and denoted by

$$e_i = y_i - \hat{y}_i \quad (i = 1, 2, \ldots, n).$$

Since OLSE does not give the estimation of σ^2, and by employing maximum likelihood ratio principle, we have the maximum likelihood estimation of β is still $\hat{\beta}$, and in the same time, we get maximum likelihood estimation (MLE) of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_i x_i)]^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2.$$

However this estimation is biased estimation of σ^2, and usually in the real application unbiased estimation of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_i x_i)]^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2.$$

It is worth to mentioning here the above maximum likelihood estimation is obtained under normal distribution assumption of $\epsilon\square N(0,\sigma^2)$, while the OLSE doesn't need such assumption.

2.2. Statistical test

(*Significance test of regression equation*) The significance test for linear regression equation if F text, and F text is based on sum of squares decomposition to text the significance of regression equation from regression effect directly. The sum of squares decomposition is

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where $\sum\limits_{i=1}^{n}(y_i - \bar{y})^2$ is called sum of squares of deviations (SST), which reflects the fluctuation degree of y, after the linear regression equation is established, it can be decomposed into sum of squares of regression and sum of squares of errors. $\sum\limits_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ is called sum of squares of regression (SSR), it is obtained by regression equation, i.e., fluctuation of x. $\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2$ is called sum of squares of errors (SSE), it reflects the fluctuation which cannot be explained by x, it is caused by the factors out of control of x. Hence the sum of squares decomposition formula can be in short given by

$$SST = SSR + SSE.$$

From the above, we know that the bigger sum of squares of regression (SSR) is, the better effect of regression is, and based on this fact we can formulate F test statistic $F = \frac{SSR/1}{SSE/(n-2)}$.

Under the normal distribution assumption, if the original assumption $H_0 : \beta_1 = 0$ holds, then $F\square F(1, n-2)$. If the value F is bigger than critical value $F_\alpha(1, n-2)$, the original assumption is penalized, which means regression equation is significant, that is, x and y have significantly linear relationship.

(*Significance test of regression coefficient*) In regression analysis, the significance test of regression coefficient is to text the significance of x and y, and usually we use t-text to check the significance of regression coefficient. Its test statistic is $t = \frac{\hat{\beta}_i}{\sqrt{SSE/(n-2)}}$. If the original assumption $H_0 : \beta_1 = 0$ holds, then $t\square t(n-2)$. If $|t| \geq t_{\alpha/2}$, the original assumption is penalized, which means β_1 is not zero significantly, that is, y is unary linear regression with respect to x. Otherwise, the original assumption is not penalized, which means that it does not hold that y is unary linear regression with respect to x.

2.3. Applications of regression diagnostic and regression model

(*residual analysis*) Determining whether the selected regression model can fit the data we studied is called *regression diagnostic*. There are many methods to check the regression model, among them, the most important method is residual analysis and collinear diagnostic. This paper focuses on residual analysis.

Residual analysis is the method by which the reliability, periodicity, or other interference of the data is analyzed based on the information provided by the residuals, and the method is used to analyze the correction of the assumptions of the model. The so-called residual is the difference between the observed value and the regression fit value, the residuals e_i can also be regarded as estimates of the error term ϵ_i. The contents of the residual analysis include the feasibility of the linear regression equation, the hypothesis of the variance of the error term, the hypothesis of independence and the distribution of the normal distribution, and whether there is an anomaly in the observed values.

The text methods of residual analysis are: frequency test, QQ test, residual map test. We usually do residual analysis by using residual map, it checks visually the variance and independence of the error term, and whether the anomaly value is included in the model. The scatter plot, with self-variable x, dependent variable y or regression fitting value $\hat{y}$ as the horizontal axis, with the residual as the vertical axis, is called *residual map*. In general, if a regression model satisfies the basic given assumptions, all residuals should vary randomly in the vicinity of $e = 0$, and in a region where the magnitude of the change is not significant, the region is usually $(-2, 2)$.

The purpose for building a regression model is to put it into practice, and the most important application is forecasting. When we use regression equations to predict the value of dependent variables, there are point forecast and interval forecast, respectively.

(*Point forecast*) Point forecast is to specify a single point value as the forecast value of new dependent variable. Suppose the "optimal" regression equation that has passed the examination is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p.$$

When the new observed value of independent variable is $x = x_0$ and $p = 1$, the corresponding dependent variable is $\hat{y_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

(*Interval forecast*) There are two cases for interval forecast of independent variable, one of which is to forecast the new value of the dependent variable and the other is to forecast the average among new values of the dependent variable. We mainly discuss the the first case here.

2.4. Correlation coefficient

The correlation coefficient is a measure that determines the degree to which two variables' movements are associated. If the correlation coefficient is computed according to the whole sample data, it is called sample correlation coefficient, which is also simply called correlation coefficient (r) and is expressed as

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}}.$$

The correlation coefficient r has the following properties:
1. r satisfies that $-1 \leq r \leq 1$. If $0 < r \leq 1$, it shows that the independent variable x and the dependent variable y are of positive linear correlation. If $-1 \leq r < 0$, it shows that the independent variable x and the dependent variable y are of negative linear correlation.
2. If $r = 0$, then it shows that there is no linear correlativity between these two variables. If $r = 1$, it shows that there is a perfect positive correlation between these two variables, and if $r = -1$, it shows a perfect negative correlation.
3. When $-1 < r < 1$, to explain the degree to which two variables' movements are associated, the degree of correlation is usually divided into the following cases:

$$\begin{cases} |r| \geq 0.8, \text{ high correlation} \\ 0.5 \leq |r| < 0.8, \text{ moderate correlation} \\ 0.3 \leq |r| < 0.5, \text{ low correlation} \\ |r| \leq 0.3, \text{ weak correlation or no correlation.} \end{cases}$$

This explanation is based on the tests of significance for correlation coefficient.
4. Note that there may be no causality between variables even the correlation coefficient is large. There is a possibility that the two variables are both influenced by another variable and hence has a strong correlation. Moreover, when correlation coefficient is very close to 0, it just shows that there is no linear correlation between the variables, but some nonlinear correlation may exist. Sometimes, some extreme data can also effect the correlation coefficient.

3. Analysis of relationship between college entrance examination scores and university achievement

3.1. Data preprocessing

This paper collected the college entrance examination results and university mathematical scores of grade 2016 freshmen from *College of Engineering*, *College of Economics and Management*, and *College of Science* in Yanbian University. Since the freshman studied just the first semester, we only use the mathematical scores of first semester to taking our analysis. Moreover, since Yanbian University accepted students throughout the whole country, the college entrance examination results measurement standards are not unified, therefore we need to deal with pre-processing for the original data, that is, deleting some useless data.

Data deletion mainly focuses on student sample deletion, because the calculation of college entrance examination scores in Jiangsu Province and other provinces are inconsistent, hence we delete the source of students in Jiangsu Province. Also, we delete such student samples who did not participate in the final examination due to certain reasons. After the deletion of the original score data, totally 1020 valid data were left. As shown in Table 1: (Without the students in Jiangsu Province and the students who did not take the final exam!)

Table 1. Data preprocessed

Name	Gender	Nation	Total	Province	College	SATS	SR	TS
Q. Ai	F	Han	525	Tianji	Engineering	123	50	63
H. An	M	Han	523	Gansu	Engineering	93	89	92
J. An	M	Chaoxian	459	Jilin	Engineering	90	63	74
M. An	F	Chaoxian	399	Jilin	Science	63	46	60
T. An	M	Chaoxian	460	Jilin	Engineering	99	38	49
X. An	M	Chaoxian	557	Jilin	Engineering	126	40	53
Q. Ao	F	Man	527	Jilin	Engineering	112	65	71
G.M. Bai	M	Chaoxian	457	Heilongjiang	Economics	50	3	30
H. Bai	M	Chaoxian	420	Jilin	Science	80	56	69
X. Bai	F	Han	551	Jilin	Economics	125	77	84
...	...	...	...	...	...	...	...	...
SATS: SAT Score; *SR*: Score Results; *TS*: Total Scores.								

We can not directly compare the student scores, for the reason that how strict the scoring teacher is, how different the student's learning ability is and some other factors which may affect the student's scores. Hence this paper uses the *relative rank*(i') to reflect the student's performance level, and then to analyze the correlation between college entrance examination results and university achievements, defined by:

$$i' = 1 - \frac{i-1}{n},$$

where n denotes the number of students who taking the final exam, after descending the order of the test scores, the score of student x ranked by i, the score higher than student x ranked by $i-1$, and the same sub-data has the same ranking (see Table 2).

TABLE 2. Data after calculating relative ranking

Name		Total	Ranking	R.R		SATS	SR	TS	Ranking	R.R
Q. Ai	···	525	582	0.43	···	123	50	63	710	0.30
H. An	···	523	599	0.41	···	93	89	92	126	0.88
J. An	···	459	816	0.20	···	90	63	74	512	0.50
M. An	···	399	974	0.05	···	63	46	60	776	0.24
T. An	···	460	815	0.20	···	99	38	49	854	0.16
X. An	···	557	236	0.77	···	126	40	53	845	0.17
Q. Ao	···	527	566	0.45	···	112	65	71	482	0.53
G.M. Bai	···	457	820	0.20	···	50	3	30	1016	0.00
H. Bai	···	420	933	0.09	···	80	56	69	626	0.39
X. Bai	···	551	295	0.71	···	125	77	84	320	0.69
···	···	···	···	···	···	···	···	···	···	···
R.R: Relative Ranking; *SATS*: SAT Score; *SR*: Score Results; *TS*: Total Scores.										

3.2. Analysis on the correlation between college entrance examination and college mathematical achivements

3.2.1. *Professional analysis.* First of all, using *SAS* software on the total score of the college entrance examination and the relative ranking of university mathematical scores to calculate the *correlation coefficient*, as shown in Table 3. (Note that in Table 3, A: Geography; B: Electronic Information Engineering; C: Business Management; D: Chemistry; E: Chemical Engineering and Technology; F: Mechanical Engineering; G: Economics; H: Civil Engineering; I: Physics)

TABLE 3. The RC of the R.R of each SATS and the R.R of university achievement

	A	B	C	D	E	F	G	H	I
RC	0.328	0.439	0.358	0.472	0.508	0.535	0.376	0.716	0.381
RC: Relative Coefficient; *R.R*: Relative Ranking; *SATS*: SAT Score									

From Table 3, we know that correlation coefficient of Civil Engineering, Mechanical Engineering, and Chemical Engineering and Technology is a little high, that is, the college mathematical achievements of the students in *College of Engineering* has high correlation with college entrance examination results. This is because *College of Engineering* of Yanbian University is the predecessor of *Department of Science and Technology* of Yanbian University, it showed that *College of Engineering* and *College of Science* has a close relationship. Moreover, most of the principles involved in computer programming are based on mathematical theory, and it can be said that solid mathematical foundation is an important theoretical basis for computer programming. Therefore, the college entrance examination of the students in *College of Engineering* has great influence on college mathematical achievements.

3.2.2. *Sub-regional analysis.* Similar to the above method, now we give correlation coefficient between the college entrance examination results and college mathematical achievements of the students from different regions, the results are shown in Table 4.

TABLE 4. RC of RR of college entrance examination in each region and RR of university achievement

	North East	North	Central	East	South	North West	South West
CC	0.388	0.097	0.231	0.343	-0.030	0.248	0.177
RC: Relative Coefficient; *RR*: Relative Rank; *CC*: Correlation Coefficient							

In which, the Chinese regional classification criteria are as follows:

East China: Shandong, Anhui, Shanghai, Zhejiang, Fujian; Central China: Henan, Hubei, Hunan; North China: Tianjin, Hebei, Shanxi, Inner Mongolia; North East: Heilongjiang, Jilin, Liaoning; North West: Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang; South China: Guangdong, Guangxi, Hainan; South West: Sichuan, Guizhou, Yunnan, Chongqing, Tibet.

From the above results, we know that the correlation coefficient between the college entrance examination results and university scores of North East is the biggest, and then the East China. There are two reasons for this result: on the one hand, the number (523 persons) of students from North East is highest, and then the East China (116 persons), that is, the sample size from North East and East China is bigger than other regions, hence proportion in the correlation coefficient analysis is relatively large. On the other hand, Yanbian University is located in Yanji City, Jilin Province, belonging to the North East. Therefore, students from North East are more quickly to adapt to the teaching environment of the university than other areas. Moreover, in the classroom, they are connected with those of high school. This will affect the efficiency of learning and the quality of learning, and affect the results of final exam. Therefore, the correlation between college entrance examination results and college scores of the students from North East is more relevant.

3.3. The total score of college entrance examination and the regression analysis of college mathematical achievements

3.3.1. *The significance of the location of the students on the university of the significance of the test.* (I) *research methods*:

The original data are divided into two samples (the south and the north) according to the students original locations. The two-sample mean significance test method is used to judge whether there is a significant difference between the university scores of the students in the South and the North. The number of students from the South is 265, and the number of students from the North is 755.

It is hereby declared that the provinces in the southern and northern provinces are as follows:

South of China: South China (Hainan, Guangdong, Guangxi), Yunnan, Guizhou, Fujian, Anhui, Hunan, Hubei, Sichuan, Chongqing, Shanghai and Zhejiang.

North of China: Heilongjiang, Jilin, Liaoning, Inner Mongolia, Tianjin, Hebei, Henan, Shandong, Xinjiang, Tibet, Gansu, Qinghai, Ningxia, Shaanxi, Shanxi.

(II) *research results*:

Table 5. Two statistical tables of samples

Place	No.of People	Aver. Score	Stan. Dev.	Mean Stan. Err.
South	265	77.69057	15.26094	0.93747
North	755	69.55099	17.99148	0.65478

From Table 5, we can see that the average score of university students from South and North is 77.70 points and 69.56 points, respectively.

It is known from Table 6 that the probability of the F test is 0.0016, which means that the variance test results of these two samples are non-homogeneous.

At this time, the significance probability of t test is < 0.0001 (less than 0.05), which shows that the difference of university achievement between two types of students is significant, that is, the achievements of students from South are obviously higher than those of North.

(III) *results analysis*:

Many different reasons caused significant differences of college achievements between students from South and North. In view of the college entrance examination results, the average score (556.97) of students from South is higher

Table 6. Results of 1022 college students' test

	VH test		t-test of the mean of two independent samples				
	F-s	P-v	t-s	FD	P-v(d)	The 95% CI of difference	
						UCI	LCI
VH	1.39	0.0016	-6.580	1018	<0.0001	1.1341	1.687
VNH			-7.118	539.45	<0.0001	-10.57	-5.71

F-s: F-statistics; *P-v*: P-value; *t-s*: t-statistics; *FD*: Freedom Degree; *P-v(d)*: P-value double; *CI*: Confidence Interval; *UCI*: Upper Confidence Interval; *LCI*: Lower Confidence Interval;
VH: Variance Homogeneous; *VNH*: Variance Non-Homogeneous.

than the college entrance examination of the average score (503.82) of students from North. In view of learning attitude, students from South are usually more diligent and harder than students from North. Comparing to North, most of the southern region faces to the pressure of further education, competition is more intense, this is one reason why students from South should have a strong self-learning ability. And for college students, self-learning ability is essential on academic performance.

3.3.2. *Significant test of the impact of gender on university performance.* (I) *research methods*:

In the same way as above, the university scores are divided into two independent samples according to their gender, and the significance of the two samples was used to judge whether there were significant differences between male and female college students' achievements, here are 522 boys, 498 girls.

(II) *research results*:

Table 7. Two statistical tables of samples

Gender	No. of People	Aver. Score	Stan. Dev.	Mean Stan. Err.
Male	522	67.98084	18.22728	0.79779
Female	498	75.52811	16.23111	0.71733

From Table 7, we know the average score and standard deviation of male and female college achievements.

As it is shown in Table 8, the significance probability of the F test is 0.0091 (less than 0.05), which means that the above two samples of the university results of the variance test results are non-homogeneous.

At this time, the significance probability of t test is < 0.0001 (less than 0.05), which indicates that the difference of university achievement between two

TABLE 8. Results of 1022 college students' test

	VH test		t-test of the mean of two independent samples				
	F-s	P-v	t-s	FD	P-v(d)	The 95% CI of difference	
						UCI	LCI
VH	1.26	0.0091	-6.972	1018	<0.0001	-9.67	-5.42
VNH			-6.991	1014	<0.0001	1.0595	1.5005

F-s: F-statistics; *P-v*: P-value; *t-s*: t-statistics; *FD*: Freedom Degree; *P-v(d)*: P-value double; *CI*: Confidence Interval; *UCI*: Upper Confidence Interval; *LCI*: Lower Confidence Interval;
VH: Variance Homogeneous; *VNH*: Variance Non-Homogeneous.

types of students is significant, that is, the students' university achievement is influenced by gender.

(III) *results analysis*:

The average score of male and female college students is 67.98084 points and 75.52811 points, respectively. The results showed that gender had a significant effect on college students' university achievement. This is because the boys prefer to move rather than girls, and the girls are careful and cautious, and also girls know how to study through the acquired efforts, cultivate self-learning ability. Therefore, these reasons caused such phenomenon that university achievement of girls is better than boys.

3.3.3. *Regression analysis on college achievements and college entrance examination scores of college students.* (I) *research methods*:

Firstly, college achievements and college entrance examination scores of college students are used as independent variables and dependent variables, and investigate whether there is a linear relationship between the variables, and then the regression model is established by unary linear regression analysis.

(II) *research results*:

TABLE 9. Regression model of the summary table

Model	CC	JF	JF of the correction	Square root of MSE
Contains constant items	0.36908	0.9470	0.9470	16.9790
No a constant item		0.1362	0.1354	0.92986
CC: Correlation Coefficient; *JF*: Judgment Factor; *MSE*: Mean Square Error.				

Table 9 shows that the correlation coefficient between college student achievement (Y) and college entrance examination (X) is 0.36908, the changing of Y is

TABLE 10. Variance analysis table of regression equation

Model	Sum of square	FD	The square of the mean	F statistic	p value
Regression	5262756.17	1	5262756.17	18255.4	<.0001
Residual	294338.830	1021	288.2848		
Total Distance	5557095.00	1022			
FD: Freedom Degrees.					

TABLE 11. Regression coefficient estimation

Model	Non-standardized regression coefficients				
	Estimated Value	LCI	UCI	t statistic	p value
SATS	0.6719	0.6621	0.6816	135.11	<0.0001
SATS: SAT Score; *UCI*: Upper Confidence Interval; *LCI*: Lower Confidence Interval.					

FIGURE 1. The residual graph of college entrance examination and student's residual

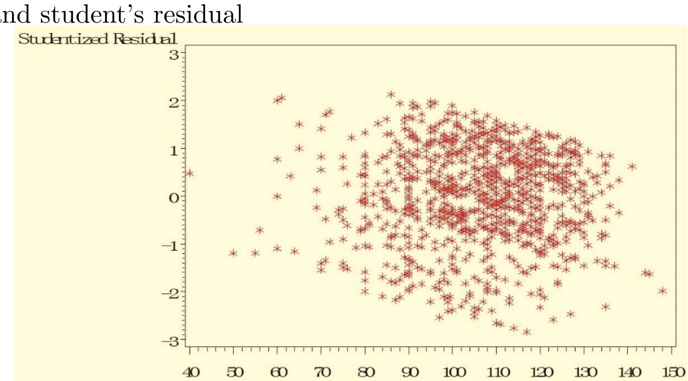

TABLE 12. Normality test of student's residual

Position Text: $Mu0 = 0$				
Test	Statistics		P-value	
Student t	t	0.087176	$Pr > \|t\|$	0.9305
Signal	M	30	$Pr \geqq \|M\|$	0.0646
Signal Rank	S	12582	$Pr \geqq \|S\|$	0.1814

determined by decision X in the size of 94.7%, and by other stochastic factors in the size of 5.3%.

From Table 10, we know the probability of significance is < 0.0001 (less than 0.05), therefore there is a linear relationship between university achievement and college entrance examination, from Table 11, we get the regression equation is:

$$Y \text{ (university) } = 0.6719X \text{ (college entrance examination results)}$$

Meantime, the significant probability of regression coefficient is < 0.0001, which means that the regression coefficient is significant.

Although the regression equation passes the t-test and the F-test, it can only show that the linear relationship between the variables X and Y is significant, but it does not guarantee that the data fits well. Therefore, before using the regression model to predict, the regression model should be tested. According to the data, calculating the residual e, and then obtaining studentized residual SRE, in the sequel making the residual map by taking the independent variable X as the horizontal axis, the studentized residual SRE as the vertical axis. From Figure 1, we can see that all points fluctuate randomly in the range $[-3, 3]$, without any rules. It shows that the regression model has no outliers and satisfies the assumption of variance homogeneity. As it can be seen from Table 12, the studentized residual SRE satisfies the approximate normal distribution, that is, the residuals pass the normality test, thus completing the residual analysis.

In addition, it can be seen from Table 9 that the decision coefficient (0.1362) of the regression equation with constant terms is much smaller than the decision coefficient (0.9470) of the regression equation without the constant term. Therefore, we use a regression equation without constant term to estimate the student's university achievement is more reasonable. In other words, if a student's college entrance examination score is 100 points, then his university achievement may be 67.19 points, and the confidence interval of 95% university achievement is (66.21, 68.16).

(*III*) *results analysis*:

From the above results, we can see that the correlation coefficient is 0.36908, which means that the college entrance examination results on the impact of university performance is not so big. This is caused by the following two reasons. On the one hand, we emphasis on the systematic knowledge in college mathematics teaching, which is difficult to reconcile with the expansion of enrollment, large class teaching model and the differences of students' original place.

The amount of information taught in the classroom is also very limited, and the content of the lectures is too low in connection with the content of high school, even in the class of high school as a "top student of mathematics" may not be able to study very smoothly and well in college class. On the other hand, most of the students in high school are overly dependent on the teacher, and focusing on learning the content of the test outline taught by the high school teacher, as long as the understanding of several typical examples of problem-solving methods will be able to get better results. While the college study pays more attention to students' self-learning ability, initiative and suitable for their own learning methods.

4. Conclusions

Through the correlation analysis and regression analysis of college entrance examination results and college achievements of grade 2016 freshmen from *College of Engineering*, *College of Economics and Management*, and *College of Science* in Yanbian University. We have the following conclusions:

(i) We took the majors and locations of students as the starting point to do the relevant analysis, and we found that college entrance examination results and college achievements have certain relevance. Among them, the results of the correlation analysis show that the college students' college entrance examination scores and university mathematics scores are relevant more in some sense; according to the regional correlation analysis, we found that college entrance examination scores and college achievements of students from North East and East China students are relevant more in some sense.

(ii) The sample was divided into two parts by gender and original location, and the significance of the two samples was tested. It was found that there were significant differences between the different gender and the different original location. That is, the college achievement of students from South was significantly higher than the one of students from North, and girls were significantly higher than boys.

(iii) Taking the college entrance examination scores as the independent variables, the university mathematics scores were used as the dependent variables to carry on the regression analysis, and the regression model Y *(university achievement)* $= 0.6719X$ *(college entrance examination scores)* was established. We got that if a student whose college entrance examination score is 100 points, then his university score may be 67.19 points.

Acknowledgements The authors would like to express their sincere thanks to anonymous referees for helpful and valuable suggestions and comments for the paper.

References

[1] V.V. Busato, F.J. Prins, J.J. Elshout and C. Hamaker, *Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education.* Personality and Individual Differences, Vol.29, Issue 6, pp.1057-1068 (2000)

[2] J. Ding, *Comparison of Chinese and American University Education.* Mathematical Culture. Vol.1(3), pp.80-86 (2010) (in Chinese)

[3] S. Ding, B. Miao and D. Ye, *An Analysis of the Correlation between College Entrance Examination and University Achievement*. China University Teaching, pp.29-31 (2008(11)) (in Chinese)

[4] A. Duff, E. Boyle, K. Dunleavy and J. Ferguson *The relationship between personality, approach to learning and academic performance*. Personality and Individual Differences, Vol.36, Issue 8, pp.1907-1920 (2004)

[5] J.R. Ferrari and J.T. Parker, *High School Achievement, Self-Efficacy, and Locus of Control as Predictors of Freshman Academic Performance*. Psychological Reports, Vol.71, Issue 2, pp.515-518 (1992)

[6] L. Nie, *Current Situation of College Mathematics Teaching Quality and Countermeasures*, Journal of Capital University of Economics and Business. Vol.16(6), pp.122-124 (2014) (in Chinese)

[7] M.Y. Xiang and L. Sun, *An Exploration and Practice of University Mathematics Teaching Research Methods*. Journal of Huangshan University, pp.82-85 (2016(5)) (in Chinese)

[8] B. Yu, Y. Xia and Y. Zhang, *An Analysis of the Weak Correlations between Mathematical Achievements and High Achievements in College Entrance Examination - A Case Study Based on A University*. Journal of Chuzhou University, Vol.13(5), pp.15-17 (2011) (in Chinese)

[9] H. Zhang and L. Zhang, *Exploration on the Teaching Method of Basic Mathematics Course Based on "Excellence Plan"*, Journal of Tianjin Chengjian University, Vol.20(2), pp.146-148 (2014) (in Chinese)

Guang-Ri Piao
Department of Mathematics, Yanbian University, Yanji, China
E-mail address: grpiao@ybu.edu.cn

Mei-Hong Cui
Department of Statistics, Yanbian University, Yanji, China.
E-mail address: 1531146727@qq.com