

A study on the relationship between students' course scores and starting wages based on principal component regression

Yi Yang^{1*}, Yan Song²

^{1,2}Department of Information Technology, Hunan Women's University,
Changsha, China

*Corresponding author e-mail: snryou@126.com

Keywords: course score, starting wage, principal component regression analysis.

Abstract. To accurately evaluate students' professional ability, optimize their curriculum, principal component regression analysis method is used to design the principal component regression model about the relationship between course scores and starting wages. Some empirical analysis is made based on score data of students which majored computer science and technology in Hunan women's university. The experiment result shows that model can more accurately describe the relationship between course scores and starting wages, which can be used as a tool to evaluate students' professional ability, forecast analysis on the employment situation for students.

1 Introduction

Solving the problem of college students' employment is not only a realistic consideration to alleviate the current employment pressure of college graduates, but also a forward-looking strategic linked with deepening the reform of higher education and promoting the national competitiveness [1]. Many studies have shown that one of the most important reasons for difficulty in obtaining employment and low salary for college graduates is their lack of employment ability. To enhance the employ ability of students for successful employment and higher starting wage, we first have to find some methods to evaluate students' professional ability and the relationship between students' professional ability and social employment ability. For college students, their course scores can reflect their professional knowledge and professional technical application ability during college study. While starting wages represent of specialization degree of graduate students after they work. So the problem of the relationship between professional ability and employment ability can be transformed into a problem of the relationship between course scores and starting wages.

In this paper our objective includes three sides: the first is to develop a model framework to predict starting wage level by studying the relationship between course scores and starting wages; the second is to use it to reflect the level of college education and the last one is to provide a reference method for employers to select graduates and provide students with professional ability training target.

2 Literature Survey

At present, there are more articles about how to make a comprehensive evaluation and scientific analysis about students' course scores, which are usually carried out by using principal component analysis or data mining techniques. For example, Jin [2] used the factor analysis method to evaluate student course scores and showed professional ability hidden behind student course scores. Yu [3], Wan [4], He [5] adopted multivariate linear regression and data mining analysis method to correlation analysis and prediction on student achievement. On the other hand, some scholars have studied definition, training and evaluation of college students employ ability. For example, Jia [1] et al. analyzed the structure of employ ability of college students. Zhu [6] et al. use behavioral event interview and survey method to study the factors affecting employers' choice and graduates' work, their cognitive differences about employment ability effect. Chen [7] made comparison analysis between Chinese college student and USA peer on college curriculum.

Although the above papers discussed about some college students' ability problem, such as: evaluation of course score, the abilities constitution of students and so on, but did not explored the correlation and influence between professional ability during school study(course scores) and students' professional ability during the period of their work (starting wages).Therefore, the above articles are difficult to achieve accurate evaluation of students' professional ability and enhance employment of students.

3 Proposed Method

Regression analysis is a method to predict dependent variables using one or more independent variables. In this paper, course score is regarded as the independent variable, while starting wage is taken as the dependent variable. The starting wage can be expressed as a linear combination of the scores of all subjects.

In the establishment of multivariate regression model about course scores and starting wages relationship, in order to reflect the influence of the starting wage information, it often uses as much as possible parameters including employment related scores of all courses in the model , but there are often collinarity between course scores, namely the existence of information overlap problem [7]. The number of courses is too large, which will not only increase difficulty of modeling, but also affect estimation of model parameters and have a negative impact on model prediction.

In order to satisfy the rationality of the established model and improve prediction accuracy of the model fitting, in the paper, principal component analysis (PCA) is applied to reduce the number of independent variables (course scores) by linear transformation, so as to reduce the difficulty of modeling and improve the accuracy of fitting coefficients.

4 Experimental Results

4.1 Data and Materials

Student score data is from Hunan Women's University 2015 graduate score data which consists of 86 course scores from computer science and technology professional 74 graduate students. The starting wages are also from the 74 graduates' wages during their probationary period. After data clean, 70 graduating students are picked out. In the process of model building, the data of 50 students are selected as training data, and the data of the remaining 20 students are used as test data. The general situation of training data is shown in Table 1 .

Table.1 Train data summary (n=50)

ITEM NAME	$\bar{d} \pm s$	ITEM NAME	$\bar{d} \pm s$
X1: DATA STRUCTURE	82.51±10.33	X6: WEB APPLICATION TECHNOLOGY	81.14±13.33
X2: LARGE SCALE SOFTWARE SIMULATION TRAINING	83.97±5.97	X7: ANIMATION TECHNIQUE	79.17±5.07
X3: OBJECT-ORIENTED PROGRAMMING	80.54±7.56	X8: ENTERPRISE LEVEL PROJECT SIMULATION TRAINING	83.10±5.71
X4: SOFTWARE TEST	81.77±5.78	X9: GRAPHIC ARTIST DESIGNER	87.84±3.63
X5: DATABASE PRINCIPLE	84.60±7.78	Y: STARTING WAGE	4035.70±910.2

4.2 methods

Principal component analysis of course score data is conducted based on the cumulative contribution rate of the initial eigenvalue greater than 85%. In order to determine the number of principal components, correlation coefficient matrix between the course scores is first calculated. Then Scree test is performed according to coefficient matrix and the characteristic value is greater than 1. The result is shown in Figure 1.

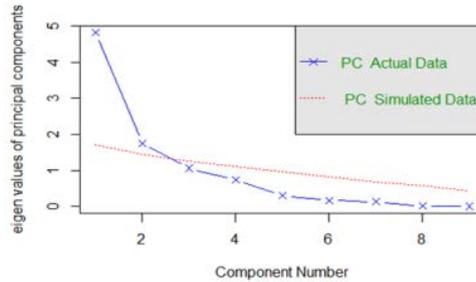


Fig.1 Scree Test Plot about the factors of impact on employment wages

According to the result of the Scree test, eigenvalues of the first 3 principal components are larger. As shown in Table 2, cumulative contribution rate of the first 3 principal component eigenvalues has reached 85.1%, which is greater than cumulative

Table.2 Principal Component eigenvalue and contribution rate

PC	INITIAL EIGENVALUE			EXTRACTION SUMS OF SQUARED LOADING		
	EIGENVALUES	VARIANCE (100%)	CUMULATIVE VARIANCE (100%)	EIGENVALUES	VARIANCE (100%)	CUMULATIVE VARIANCE (100%)
PC 1	4.820	53.60	53.60	4.820	53.60	53.60
PC 2	1.755	19.60	73.20	1.755	19.60	73.20
PC 3	1.060	11.90	85.10	1.060	11.90	85.10

contribution rate of initial extracted eigenvalues 85%. We can construct multivariate regression analysis expressions of the model as Equation 1:

$$\begin{cases}
 PC1=0.9X1+0.82X2+0.80X3+0.78X4+0.76X5+0.76X6+0.54X7+0.41X8-0.19X9 \\
 PC2=-0.3X1-0.04X2-0.16X3+0.32X4+0.47X5-0.57X6+0.63X7+0.06X8+0.76X9 \\
 PC3=0.21X1+0.34X2-0.36X3+0.49X4-0.20X5+0.19X6-0.35X7+0.55X8+0.42X9
 \end{cases} \quad (1)$$

PC1 reflects outstanding features of software engineering in the specialty curriculum; PC2 reflects some students' "weighting practice and lighting theory"; PC3 reflects some students' prefer art design. According to the three principal component regression analyses, we can get short Equation 2:

$$WAGE= 17.14PC1-20.22 PC2+32.65PC3-3296.31 \quad (2)$$

Finally, put Equation 1 into the Equation 2, prediction equation between final course scores and starting wages:

$$WAGE=26.914X1-14.844X2+28.302X3-2.66X4+5.317X5-4.885X6+22.668X7+30.605X8-2.926X9-3296.31 \tag{3}$$

Here correlation coefficients $R = 0.894$, $R^2 = 0.891$, $R^2_{adj} = 0.872$. It shows that the established principal component regression analysis model can explain the variation of starting wage by 89.4%, the error is acceptable and the fitting effect is better.



Fig. 2 The comparison between contract salary and predicted result

The data of the first set of 50 samples are put into Equation 3 to get predicted salary. Compared them with actual starting wages, and the results are shown in Figure 2. With Equation 3 to predict the latter 20 samples data, maximum absolute error of the model is 1099.21, average error is 343.46, minimum error is 14.05, while the corresponding relative error value are 0.3141, 0.0934, 0.0035. It's showed in Table 3. Therefore, the prediction error of the model can meet the demand of the forecast of actual starting wage, and the accuracy and availability of the model are verified.

Table.3 Comparison of actual and predicted values

NO	STARTING SALARY		ERROR VALUE	
	ACTUAL VALUES	PREDICTED VALUES	ABSOLUTE ERROR	RELATIVE ERROR
62	4000	4014.05	14.05	0.0035
66	3500	4599.21	1099.21	0.3141
AVERAGE	3675	3663.11	343.46	0.0934

5 Conclusion

In the paper, an idea is implemented to build the relationship model between course scores and starting wages. The effect of course scores on starting wages can be abstracted into several main aspects according to principal component analysis. Each aspect reflects a certain professional ability of students at school, but also determines the direction of the future students can engage in professional work after they graduate. Based on the students' school course scores, the model can be used to predict the employment situation of students. Addition of more students' course scores as well as more starting wages and implementation of the same model for prediction of more starting wages can be helpful in college education system.

Acknowledgement

The research is funded by the "twelfth five-year" Education Science Plan Project of Hunan province of China (Grant Nos.XJK014AXX002), the scientific research project of Education Department of Hunan Province of China (Grant Nos.15C0696).

References

- [1] Jia Lijun and Guan Jingjuan, "Structure of College Students' Employability", *Research in Educational Development*, 2013(13), PP: 51-53.
- [2] Jin Shuying, Zhang Huixue, "Enlightenment of Performance Evaluation Based on Factor Analysis to Interdisciplinary Talents Cultivation", *Journal of Guizhou Normal University (Social Sciences)*, 2015(3), PP: 151-154.
- [3] Yu nengfu, "Application of multivariate linear regression in analysis of the correlation of students' grades", *College mathematics*, 2007(4), PP:44-46
- [4] Wan xinghuo, Zheng junling, Jin yongchao, "A prediction model on the performance of college students based on Kernel Principal Component Analysis", *Mathematics theory and application*, 2013(4),PP:61-64
- [5] He chu, Song jian, Zhuo tong, "Curriculum association model and student performance prediction based on spectral clustering of frequent pattern", *Application Research of Computers*, 2015(4), PP:2930-2934
- [6] Chen Juntu, "College Students' Employment and University Curriculum", *China Higher Education Research*,2012(3),PP:75-78
- [7] Zhu Guowei and Huang Jun, "A Study on Influencing Factors of the College Student's Employability", *Educational Research*, 2011(8),PP: 66-68
- [8] Ding meifang, "PCA application in students' score analysis", *Journal of Technology College Education*, 2003(5),PP:71-74