

An allocation model of educational finance based on Big Data

ChaoQun Sheng^{1, a}

¹North China Electric Power University, Hebei 071000, China;

^a327417065@qq.com

Keywords: data screening, different degrees, discrete data statistics

Abstract. This paper attempts to provide a thorough allocation model of educational finance for Goodgrant Foundation, with different kinds of factors taken into consider. In order to fulfill the optimum allocation of funds and pin down a favorable investment strategy, we build two models to research into school choice, fund distribution, return prediction, investment duration, which is a quite influencing factor, and other issues. With regard to candidate schools and non-candidate schools, we choose methods in statistics such as data screening, different degrees comparison, discrete data statistics and principal component analysis, etc, and pick out the important indicators to differentiate the candidate schools from non-candidate schools. In addition, on the basis of analyzing these indicators, we introduce the concept of “improvement factor”, with a purpose of making sure the rate of return model can carry through continuous and perennial return forecast and realizing the determination of investment time. On this basis, we also take time variable into full account. We transform the continuous investment time into n times investment problem with one year as a time unit, employ circular analysis to make n times allocation in accordance with the ROI maximum principle. Retention time has gained by statistics is the time duration of fund. Eventually we obtain a project about schools’ different investment time.

1. Introduction

With the advent of the era of Higher Education, Higher Education enrollment scale is increasing quickly. As the Higher Education becoming more and more general, there should be more students, more advanced equipment and higher teachers’ compensation. Unfortunately, education funds’ improvement tends to can’t catch up with the fast development of higher education. So, the shortage of funds is one of the main factors of restricting education level. It must be solved right immediately.

We are required to develop a model to determine an optimal investment strategy. It included identification of the schools, the investment amount per school, the return on that investment, and the time duration that the organization’s money should be provided. Its goal is to have the highest likelihood of producing a strong positive effect on student performance. At the same time, maximizing the function of the investment also should be taken into account. All in all, we must do so in a more reliable and scientific way instead of judging subjectively.

We are deeply convinced that the true glamour of scientific and mathematical models can help to make best use of the foundation. Aiming to do so, we will establish models that can fit large quantities of data with highest accuracy within a quick flick.

2. Model I

2.1 Index selection.

The difference in important indicators is the symbol of differences between the candidate school and other schools, this is also the basis of filtering a candidate school. After that, we believe the date list of candidate schools is absolutely not for no reason at all, it must be some significant characteristic to distinguish them from all the school database. So we contrast the two Table and attempt to find the significant difference. To some degree, the difference of these characteristic reflect some criteria how we determine the list of Candidate School. What’s more, these factors will be our important indicators to evaluate candidate schools.

2.2 Rank Schools

Make the main component analysis of the remaining group data, the following steps are as follows:

1) First, the original data were standardized, and the index values a_{ij} were converted into standardized indicators \widetilde{a}_{ij} .

$$\widetilde{a}_{ij} = \frac{a_{ij} - u_j}{S_j}, \quad i = 1, 2, 3, \dots, 2937, \quad j = 1, 2, 3, \dots, 19 \quad (1)$$

Among that,

$$u_j = \frac{1}{2937} \sum_{i=1}^{2937} a_{ij} \quad (2)$$

$$S_j = \sqrt{\frac{1}{2937-1} \sum_{i=1}^{2937} (a_{ij} - u_j)^2}, \quad j = 1, 2, 3, \dots, 19, \text{ etc} \quad (3)$$

where u_j , S_j means sample mean and sample standard deviation for the first J index. Corresponding to the ground, call $\widetilde{x}_j = \frac{x_j - u_j}{S_j}$, $j = 1, 2, \dots, 19$ standardized indicator variable.

2) Calculated correlation coefficient matrix R. correlation coefficient matrix $R = (r_{ij})_{19 \times 19}$.

$$r_{ij} = \frac{\sum_{k=1}^{19} \widetilde{a}_{ki} \widetilde{a}_{kj}}{19-1}, \quad i, j = 1, 2, \dots, 19 \quad (4)$$

$r_{ii}=1$, $r_{ij}=r_{ji}$, r_{ij} is the correlation coefficient between the i index and the j index.

3) Calculation of eigenvalues and eigenvectors.

The characteristic value of the correlation coefficient matrix R is calculated.

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{19}$, and corresponding normalized feature vectors u_1, u_2, \dots, u_{19} , $u_j = [u_{1j}, u_{2j}, \dots, u_{19j}]^T$, The feature vector is composed of 19 new index variables.

$$\begin{cases} y_1 = u_{11} \widetilde{x}_1 + u_{21} \widetilde{x}_2 + \dots + u_{191} \widetilde{x}_{19} \\ y_2 = u_{12} \widetilde{x}_1 + u_{22} \widetilde{x}_2 + \dots + u_{192} \widetilde{x}_{19} \\ \vdots \\ y_{19} = u_{119} \widetilde{x}_1 + u_{219} \widetilde{x}_2 + \dots + u_{1919} \widetilde{x}_{19} \end{cases} \quad (5)$$

Among them, y_1 is the first principal component, and y_2 is the second principal component, ... y_{19} is the nineteenth principal component.

4) Select P ($P < 19$) main components, calculate the comprehensive evaluation value.

① Calculate the contribution rate and cumulative contribution rate of the characteristic value of λ_j ($j=1, 2, 3, \dots, 19$). Call

$$b_j = \frac{\lambda_j}{\sum_{k=1}^5 \lambda_k}, \quad j = 1, 2, \dots, 19 \quad (6)$$

the information contribution rate of principal component y_j , and call

$$\alpha_p = \sum_{k=1}^p \lambda_k / \sum_{k=1}^{19} \lambda_k \quad (7)$$

the cumulative contribution rate of principal component y_1, y_2, \dots, y_p . When α_p is close to 1, then select the first p indicator variables y_1, y_2, \dots, y_p as the main component of y_p , instead of the original 19 indicator variables, so as to make a comprehensive analysis of the main components of p.

② Calculate comprehensive score:

$$Z = \sum_{j=1}^p b_j y_j \quad (8)$$

Among them: b_j is the information contribution rate of the first j principal component, and can be evaluated according to the comprehensive score value.

3. Model II

3.1 The quantization of ROI

Based on the research of above problems, with regard to the ROI, predict it directly is very difficult. As the investment not only has a direct influence on grantees, but also has an unpredictable effect on the charity or even the whole society indirectly. In this way, the quantization of ROI will provide reliable reference to the investment decisions charity.

How to quantize the ROI? For the schools, their aims are to improve the graduate rate of undergraduates and transport the talents to the society. In analysis of the data, there is significant

difference between the full-time student retention rate and part-time student retention rate. Because of the economic pressure, many college students have to have a part-time job in order to pay the high tuition. So the part-time job has been preventing these students from graduating. At the same time, the social economic contribution raised by the improvement of graduate rate, is main indicators needed to be quantized. Full-time employed graduates receive a substantial wage premium over non-graduates, on average of the order of 65 percent; that is controlling for age, experience and other characteristics usually included in such analysis, an employed graduate receives on average about 65 percent more than an employed person without a degree.[3]

Owing to the improvement of graduate after funded, calculating the additional graduates' contribution to the society in twenty years is our quantization process. Take the number of student into consideration, the same amount of investment may have different return, so we must consider the number of school student into account.

Define personal benefit on average as:

$$V_h = M/n \quad (9)$$

Where M is the total amount of investment per school, n is the number of school student, V_h is used to evaluate how large the fund is for per student.

The greater the V_h is, the more student receive the investment on average.

Similarly, in order to measure the effect of investment for the improvement of school graduation rates. We define improvement factor ϑ as:

$$\theta = kV_h/C \quad (10)$$

Where C stands for average of tuition fees. ϑ reflects the improvement of graduation rates result from investment. ϑ is proportional to the personal benefit on average V_h , and inversely proportional to Average of tuition fees C .

After define θ , calculate full-time student retention rate f_{n+1} and part-time student retention rate p_{n+1} in year $n+1$.

$$[f_{n+1} \quad p_{n+1}] = [f_n \quad p_n] \begin{bmatrix} 1+\theta & 0 \\ 0 & 1+2\theta \end{bmatrix} = [f_n(1+\theta) \quad p_n(1+2\theta)] \quad (11)$$

The return is expected to be [1]

$$R_1 = I(n \cdot \eta \cdot p_n \cdot 2\theta + n(1-\eta)f_n \cdot \theta) \times 65\% \quad (12)$$

Where η stands for part-time rate of a school, I stands for the income after ten years.

Then, Calculate the ROI of The investment plan in model I.

3.2 N years before quantization of continuous investment in research on

We consider the investment for only one year, and quantify the data of the return only for the year. On the premise of steady fund each year, school retention rates will continue to grow, change of return remains to be seen. Make an analogy with the concept of Markov [4] transition matrix, we define transfer matrix of full-time and part-time student retention rate

$$Q = \begin{bmatrix} 1+\theta & 0 \\ 0 & 1+2\theta \end{bmatrix}^t \quad (13)$$

For full-time student and part-time student retention rate after t years, we can get:

$$[f_{n+t} \quad p_{n+t}] = [f_n \quad p_n] \begin{bmatrix} 1+\theta & 0 \\ 0 & 1+2\theta \end{bmatrix}^t = [f_n(1+\theta)^t \quad p_n(1+2\theta)^t] \quad (14)$$

In the N -th year, retention rate has changed, expected return is as follows:

$$R_n = I(n \cdot \eta \cdot p_n \cdot ((1+2\theta)^t - (1+2\theta)^{t-1}) + n(1-\eta)f_n \cdot ((1+\theta)^t - (1+\theta)^{t-1})) \quad (15)$$

Having considered the influence of funding on full-time student and part-time student retention, we assume the investment of charity to each school remain constant, and we estimate the return on investment for each year of R1~R5, we get the following result:

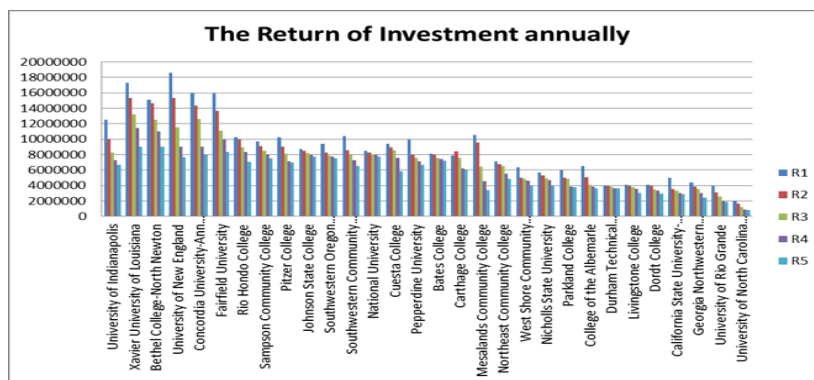


Fig 1. Return histogram

4. Summary

By observing the changes for each return value, it is not difficult to find that sustained investment will produce declining returns values, the degree of use of funds cannot be the maximum as in the first year. For students performance in the subject (a strong positive effect on student performance) is described as follows: the important indicators of performance are whether students can graduate and the social contributions after graduation, which will be our important indicators for the next optimization. We mentioned in the hypothesis, the most direct way to improve graduation rates is to improve part-time student retention rate. In the above "N-year investment quantitative calculation" also provides support for the determination of time duration, when the schools selected remain constant, what we can consider is to segment the time variable, to achieve annual maximum return in year. We get $MAX = \max_1 \&\&\max_2 \&\&\max_3 \&\&\max_4 \&\&\max_5$.

The reallocation of funds is an important factor in these problems, we need to find the right distribution in order to achieve a strong positive effect on student performance, that is the maximization of the return on investment. In addition, the continued investment time can be determined by the duration of the high return on investment.

Reference

- [1] Jeff Borland, Peter Dawkins, David Johnson and Ross Williams Returns to Investment in Higher Education [D] . Melbourne: The University of Melbourne
- [2] Lixuan Xi, The Study of the Donation to American Higher Education[D]. Nanjing: Nanjing Normal University
- [3] Zhehui Zou, A Comparative Study in Higher Education foundation of China and the United States[D]. Guangxi: Guangxi University