For office use only	Team Control Number	For office use only
T1 ________________	**52063**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2016
Mathematical Contest in Modeling (MCM/ICM) Summary Sheet

Strategies for Investing in U.S. Colleges

Summary

With the incresing importance of education, more and more foundations develop many differenct strategies to build evaluation systems for colleges and invest their money to help these colleges make contribution. This paper implement some models to generate a new evaluation system and calculate the investment amount and the time duration for each college according to this system.

First, we build a triple-metric model. After researching on the relevant filed and analyzing the provided variables, we choose the median of earnings after 10 years of graduates, the shares of the students who earn more than $250,000 per year, and the graduation rate as the main evaluation metrics. This gives us a preliminary ranking of the colleges. Then, we devise a multi-metric model by AHP (Analytic Hierarchy Process) and TOPSIS (the Technique for Order of Preferency by Similarity to Ideal Solution). We obtain three comprehensive metrics- earnings, school properties, and student statistics, and then we use the same method to obtain eight sub-metrics. We use AHP to determine the weights of different metrics in a linearization evaluation system. Then we use TOPSIS method to obtain a new ranking in a more comprehensive way.

After we obtain the ranking and the efficiency of all the colleges, we only select colleges with more than average donation, which are more convincible that these colleges are able to take advantage of the fundings. We, first, split the data set according to its pattern and density in different investment range. Then, we implement the multiple linear regression method in each data set to fit a polynomial equation and then find the optimal point, which is also the duration factor in our model. And the final total investment amount for each school is determined by the ranking score and the duration factor. After we obtain the ranking and the efficiency of all the colleges, we only select colleges with more than average donation, which are more convincible that these colleges are able to take advantage of the fundings. We, first, split the data set according to its pattern and density in different investment range. Then, we implement the multiple linear regression method in each data set to fit a polynomial equation and then find the optimal point, which is also the duration factor in our model. And the final total investment amount for each school is determined by the ranking score and the duration factor.

Finally, we discuss the strengths and weaknesses, and write the proposal letter to the CEO of Goodgrant Foundation, Mr. Alpha Chiang about the final optimal investment strategy.

Keywords: investment; foundation; charity; education; return on investment; APH; TOPSIS; DEA

1 Introduction

1.1 Defining the Problem

In recent years, the methods of school ranking and portfolio optimization have been development rapidly. In the Goodgrant Challenge problem, in order to build the models to determine the optimal investment strategy which contains the optimized 1 to N universities and the optimized donation amount and time duration, our group decided to define the problem into two parts: 1) Build models to rank universities based on students and universities performances and calculate the ROI based on the results; 2) Build models to optimize donation allocation and time duration for which high-rank university.

1.2 Data Disposal

In the first step, we select 28 relevant variables from the provided data sets. These variables can measure the performance of a college's graduates or the properties of it. Most of the data do not have complete records for all the 26 variables, so, in order to have enough data points to buid models, we decide to replace those missing values instead of simply excluding them out of the data set. According to the Carnegie Classification, we classify all the data points into 27 goups and use the group averages to replace the missing values. In the case that the group average is still missing value, we use the total average to replace it. By doing this, we believe the estimate of the missing value are accurate enough for the model construction.

The following table displays the selected variables:

Table 1: Selected Variables for Further Use

Name	Explanation
INSTNM	Institution name
UNITID	Institution ID
Type	School type
HCM2	Schools that are on Heightened Cash Monitoring 2 by the Department of Education
PPTUG_EF	Share of undergraduate, degree-/certificate-seeking students who are part-time
RET_FT4	First-time, full-time student retention rate at four-year institutions
RET_PT4	First-time, full-time student retention rate at less-than-four-year institutions
md_earn_wne_p10	Median earnings of students working and not enrolled 10 years after entry
gt_25k_p6	Share of students earning over $25,000/year (threshold earnings) 6 years after entry
Tuition07	Tuition fee in 07~08
Tuition08	Tuition fee in 08~09
Tuition09	Tuition fee in 09~10
Tuition10	Tuition fee in 10~11
Sfratio2014	Student-to-faculty ratio in 2014
SFratio2010	Student-to-faculty ration in 2010
Ftretention	Full-time retention rate
Graduation2014	Graduation rate in 2014
Graduation2010	Graduation rate in 2010
RevTuition_G	Revenues from tuition and fees per FTE (GASB)
RevTuition_F	contracts per FTE (GASB)
Donation	Revenues from private gifts, grants, and contracts per FTE (GASB)
Donation_aff	Revenues from private gifts, grants,contracts/contributions from affiliated entities per FTE (FASB)
Academic_G	Academic support expenses as a percent of total core expenses (GASB)
Academic_F	Academic support expenses as a percent of total core expenses (FASB)
Service_G	Student service expenses as a percent of total core expenses (GASB)
Service_F	Student service expenses as a percent of total core expenses (FASB)
Institutional	Institutional support expenses as a percent of total core expenses (GASB)
AdmissionYield2014	Admission yield in 2014

1.3 Model Overview

College Ranking

In order to calculate the return on investment, first we need to calculate the performance score of each school, and then compare to its donation from private funding (revenues from private gifts, grants,contracts/contributions from affiliated entities per FTE).

In the first step, we use three models to evaluate the performance of each college:

- A simple triple-metric evaluation model
- An improved multiple-metric method by using APH and TOPSIS
- A simple DEA model for efficiency evaluation

The simple triple-metric method only select three most influential variables and assign each variable to some weight. The improved method by APH and TOPSIS use the same weight system but different approaches to reach the final efficiency score. This method is more exact and less arbitrary than the first method. Finally, we used another method called DEA (Data Envelopment Analysis), which uses another kind of methodology to measure the efficieny of U.S. colleges.

Optimization

In the optimization step, we use Multiple-Linear Regression to fit models for the efficiency by using investment amount as the predictor. Then we choose the investment amount in order to maximize the efficiency line.

2 Methods

2.1 A Simple Triple-Metric Evaluation Model

2.1.1 Method Introduction

One of the most important measurements of return on investment is the earning of graduates. The better a college is, the higher the average salary of its graduatesâĂŹ is. Thus, the variables relevant to salaries are highly important to evaluate the return on investment. Another important indicator is the graduation rate. Generally, a college makes more contribution to the society with a higher graduation rate. In this simple method, we select three variables, B_1, B_2 and B_3 as the main metrices.

In the next step, we assign the weights to each variable and calculate the performance of each school. In order to simplify the model and the calculation, we first normalize these three variables and make them consistent. The original variable is denoted by $B_{original}$, the maximum value is denoted by B_{max}, and the minimum value is denoted by B_{min}. The adjusted value is:

$$\frac{B_{orginal} - B_{min}}{B_{max} - B_{min}}$$

Then we use the following equation to obtain the evaluation score, also a collegeâĂŹs performance, for each school. Here, a_1, a_2 and a_3 are the weights for B_1, B_2 and B_3, respectively.

$$Score = a_1 \times B_1 + a_2 \times B_2 + a_3 \times B_3$$

In this problem, B_1 is a variable which specifically describe the earnings of graduates of a college while B_2 and B_3, which are percentages, only describe the general feature of the graduates. Thus, we assign $a_1 = 0.4$, $a_2 = 0.25$, and $a_3 = 0.35$.

2.1.2 Result

Since the data set has many missing values, we only consider colleges with total donation greater than the average as well as calculated score larger than the average because it will be more convincible if a college's performance and donation are above average. Here, we list first twenty schools rank by the simple triple-metric model.

Table 2: Final Rank by the Simple Triple-Metric

Institution Name	Retruen on Investment
A T Still University of Health Sciences	3.30556E-05
Rosalind Franklin University of Medicine and Science	2.76433E-05
Philadelphia College of Osteopathic Medicine	2.73945E-05
Midwestern University-Glendale	2.73672E-05
University of Pennsylvania	2.71932E-05
Georgia Institute of Technology-Main Campus	2.63776E-05
Harvey Mudd College	2.62539E-05
Western University of Health Sciences	2.62444E-05
Blessing Hospital School of Medical Laboratory Technology	2.60794E-05

3 AHP(Analytic Hierarchy Process)

AHP was first developed by Tomas L. Saaty in the 1970s and has been accepted as industry standard for a long time due to its maturity . It is a wildly used decision making method for problems having multiple analyze criteria under uncertain conditions. Therefore, we decided to apply this method to this problem. [1]

3.1 Build Hierarchy Structure

The first step of AHP is to build a hierarchy structure for the problem. Generally, the structure contains three layers including goal, criteria, and alternative. The goal for this problem is to determine list of schools that are to recommended to investment. We determined three criteria based on common sense and narrowed down to total of seven alternatives according to research on related articles. The structure is shown below:

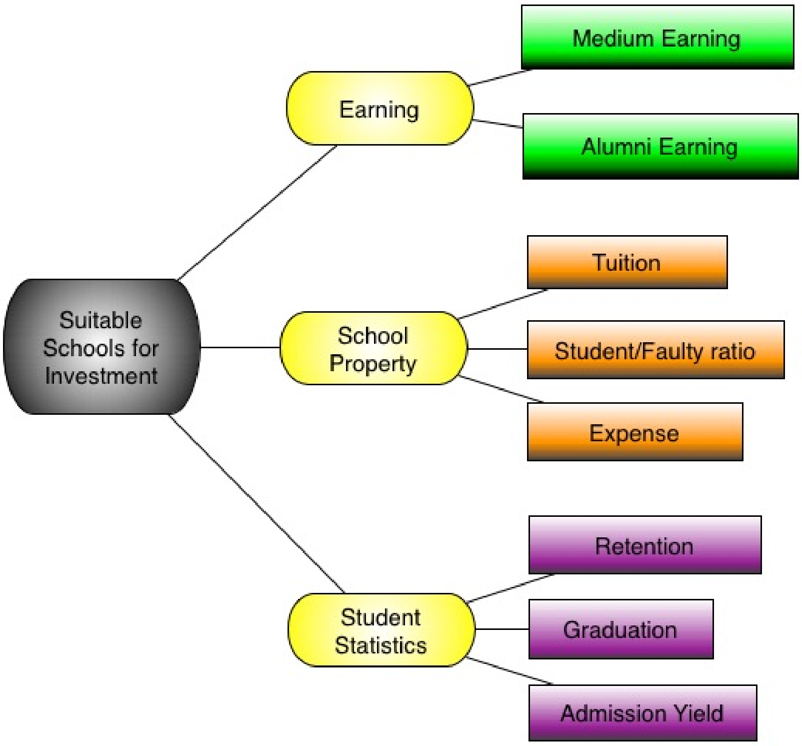

Figure 1: Hierarchy Structure

3.2 Construct Pairwise Comparison Matrix

The second step is to construct pairwise comparison matrix among each layer. Assume that a certain layer X contains n factors : $X = [x_1, x_2, ...x_n]$, then we need to rank them according to the influential of such factors to the previous layer. A nxn comparison matrix looks like the follow:

$$A = \begin{pmatrix} a_{1,3} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}$$

Here, $a_{i,j}$ represents the ith factor compares to the jth factor based on the influence to certain criteria. When $a_{i,j}$ is greater than 1, the criteria of the ith factor is better than the jth factor; when $a_{i,j}$ is equals to one, the ith and jth factor have the same importance,; when $a_{i,j}$ is smaller than 1, the criteria of the jth factor is better than the ith factor. Notice that the $a_{i,j}$ element in matrix is the reciprocal of the $a_{j,i}$ element. The standard scale for qualitatively represents the importance of certain factor to another is the below table:

Till now, we are able to construct four pairwise comparison matrix according to our hierarchy structure.

Table 3: Standard Scale for Qualitative Comparison of Alternatives

Intensity of Importance	Definition
1	Equal
3	Marginally Strong
5	Strong
7	Very Strong
9	Extremetly Strong
2,4,6,8	Intermediate values to present fuzzy input

Table 4: Four Comparison Matrix Table

Suitable Schools for Investment			
	Earning	School Property	Student Statistics
Earning	1	6	4
School Property	0.166666667	1	0.5
Student Statistics	0.25	2	1
Sum	1.416666667	9	5.5

Earning		
	Medium Earning	Alumini Earning
Medium Earning	1	5
Alumini Earning	0.2	1
Sum	1.2	6

3.3 Generate the Criteria Weight

After obtaining the comparison matrix, the next step is to finally get the weight of each alternatives. In order to do that, we need to normalize the matrix first using equation,which enables to divide each element in every column by the sum of that column:

$$q_{i,j} = \frac{a_{i,j}}{\sum_{i=1}^{n} a_{i,j}} \qquad [1]$$

After normalizing all the elements in the matrix, average each row and obtain the criteria weight $[W]$ corresponds each alternatives. Then we obtain the weight of each element in certain layer and it is obvious that the sum of the weight for elements in one layer equals to one. Since we only cares about the weight of seven alternatives, the relative weight can be calculated by multiply the weight of that alternatives to its criteria weight. The final result is as follow:

School Property			
	Tuition	Student/Faculty Ratio	Expense
Tuition	1	0.5	0.142857143
Student/Faculty Ratio	2	1	0.2
Expense	7	5	1
Sum	10	6.5	1.342857143

Student Statistic			
	Retention	Graduation	Admission Yield
Retention	1	0.333333333	1
Graduation	3	1	3
Admission Yield	1	0.333333333	1
Sum	5	1.666666667	5

Table 5: Criteria Weight of Seven Alternatives

Alternatives	[W]
Retaintion	0.0387
Medium Earning	0.5833
Alumini Earning	0.1167
Tuition	0.0101
Student/Faculty Ratio	0.0179
Graduation	0.1161
Admission Yield	0.0387
Expense	0.0786

3.4 Checking for Consistency

Checking for consistency is a really important step to make sure the criteria weight and the final result are reliable. To illustrate the consistency, let's say we have three alternatives A, B, and C. If we set the importance of A greater than B and B greater than C, then if A is smaller or equal to C, we call that inconsistent. There are four steps for checking of consistency:

Step 1
We first need to find the eigenvalue of each matrix table:

$$\lambda_{max}\text{=[sum of rows]}*[W]$$

, where the sum of rows is simply the denominator of equation [1].

Step 2
Then we need to compute the Consistency Index(CI):

$$CI = \frac{\lambda_{max} - n}{n - 1} \qquad [2]$$

, where n is the number of criteria/alternatives.

Step 3
Next, find the Random Index(RI) number from the table below. RI is only related to n, and it is always the same for the same n.

Table 6: RI Table

n	1	2	3	4	5	6	7	8	9	10	11
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51

Step 4
Finally, we calculate the Consistency Ratio(CR):

$$CR = \frac{CI}{RI} \qquad [3]$$

The result is shown below:

Table 7: CR Table

Pairwise Comparison Table Title	Consistency Ratio (CR)
Suitable Schools for Investment	0.013
Earning	0.000
School Property	0.021
School Statisitcs	0.000

According to the definition, a system is considered as consistent if $CR < 0.1$. Thus our systems successfully pass the checking of consistency.

3.5 Final Ranking

After obtaining all these information we want, we are able to rank the suitable schools that qualify the investment. For each school, simply multiply the seven criteria weight to their corresponding value under such alternatives and sum them together to get the final weight factor. Since we are interested in the return of the investment, we divide the weight factor by total donation and rank the schools according to it. The result is as follow:

Most of these TOP10 schools are medical schools, which is a reasonable result. According to research on the careers that have the best perspective, doctors and nurse are the top vocations. Therefore, investment on medical school will most likely have great returns.

Table 8: Finally School Ranking Using AHP Method

Rank	Schools	Criteria Weight	Weight/Total Donation
1	A T Still University of Health Sciences	0.744	2.88E-05
2	Rosalind Franklin University of Medicine and Science	0.652	2.46E-05
3	Philadelphia College of Osteopathic Medicine	0.558	2.29E-05
4	Midwestern University-Glendale	0.525	2.16E-05
5	University of Pennsylvania	0.348	2.00E-05
6	Western University of Health Sciences	0.519	1.98E-05
7	Midwestern University-Downers Grove	0.465	1.90E-05
8	Albany Medical College	0.728	1.90E-05
9	Washington and Lee University	0.326	1.79E-05
10	Harvey Mudd College	0.303	1.71E-05

However, there are still some weakness for using such methods. For example, it can be labor intensive while dealing with large data set; it might involves perceived loss of control due to "black box" nature of calculation; and it can become unwieldly when numbers of elements to prioritize becomes large. Therefore, we decided to use other deeper methods to solve this problem.

3.6 A Simple DEA Model for Efficiency Evaluation

We put forward to introducing Data Envelopment Analysis (DEA) model to measure and compare the efficiency of universities. [5][6][7]

3.6.1 Introduction and Description

Data Envelopment Analysis was introduced by Charnes, Cooper and Rhodes in 1978 as a nonparametric technique based on linear programming to evaluate the efficiency of organizations in the same industry.

Normally, the usual measure of efficiency, i.e,$efficiency = \frac{output}{input}$ is inadequate to rank the efficient usages of universitiesâĂŹ endowment due to the existence of multiple inputs and outputs taking into the consideration, such as graduation rate, retention rate and university graduatesâĂŹ salaries and the potential correlation among inputs and outputs. Thus, a formula based on DEA model incorporating multiple inputs and outputs for relative efficiency is used for solving the problem.

3.6.2 Model Design

In DEA, each unit used for evaluation is called Decision Making Unit (DMU). For each DMU, our group divided the selected variables into inputs and outputs. And then, build a basic evolution formula to evaluate relative efficiency for each university. Finally, we divided the relative efficiency by Total Donation amount to rank the university by the efficient usage of donation.

$$E = \frac{sum_{j\in O}u_jY_j}{sum_{i\in I}v_iXi}$$

- Set I represents the inputs of the variables, which means that, the smaller value X_i has, the better the efficiency will be.
- Set O represents the outputs of the variables, which means that, the larger value Y_j has, the better the efficiency will be.
- v_i and u_j represent the weights in set I and set O

3.6.3 The Solution to the Model

Step 1: Variable Selection and Decription After comprehensive analysis on the provided variables, our group decided to narrow the variables down to seven in DEA model.

Input attribute:

- **Total Expense:** the core expense of each university after combing GASB and FASB accounting standards of each function. Used as the input of the model because for same outcome, the lower the expense is, the higher the efficiency will be.

Output attribute:

- **Median Earning:** the median earnings of students working and not enrolled 10 years after entry. Used as the output to measure the performance of students after graduation.
- **High Salary:** Share of students earning over $25,000/year, 6 years after entry. Used as the output to measure the performance of students after graduation.
- **Retention Rate:** Percentage of students continues studying in the university after first year. Used as the output to measure the attraction of university to students.
- **Graduation Rate:** Percentage of students graduated within four years. Used as the output to measure the teaching quality of university.
- **Revenue from Tuition:** the amount of revenue comes from tuition in 2010. Used as the output to measure the resources of university.

Denominator:

- **Total Donation:** the total amount of donation received from private gifts, grants and contracts in 2010. Used as the denominator to calculate the usage of donation for each relative efficiency.

Step 2: Inemplement DEA in R program In our DEA Model, our group uses constant return to scale (CCR) which means for each DMU, the outcome is constant. We further implement our basic model into the more comprehensive model based on R program.

- **Define efficiency in the form of vectors**

$$E_k = \frac{U_k Y_k}{V_k X_k} \qquad Y_k = \begin{bmatrix} Y_1^k \\ Y_2^k \\ ... \\ Y_N^k \end{bmatrix} \qquad X_k = \begin{bmatrix} X_1^k \\ X_2^k \\ ... \\ X_N^k \end{bmatrix}$$

Y_k is the output of DMU_k, X_k is the input of DMU_k

$$U_k = \begin{bmatrix} u_1^k & u_2^k & ... & u_n^k \end{bmatrix} \qquad V_k = \begin{bmatrix} v_1^k & v_2^k & ... & v_n^k \end{bmatrix}$$

U_k is the weights of outputs , V_k is the weights of inputs

The idea behinds DEA model with CCR is to use the maximum efficiency E_k of each DMU_k to find the most effective weights of inputs and outputs. And the sum of all DMU_k âĂŸs efficiency E_k has to smaller or equal to 1. As a result, the finally implementation DEA model used by our group is

$$MaxE_k = \frac{sum_{j=1}^{n} u_j^k Y_j^k}{sum_{i=1}^{m} v_i^k X_i^k}$$

$$s.t. \frac{sum_{j=1}^{n} u_j^k Y_j^r}{sum_{i=1}^{m} v_i^k X_i^r}$$

$$u_j^k \geq \epsilon > 0, \qquad j = 1, 2, ..., n$$

$$v_i^k \geq \epsilon > 0, \qquad i = 1, 2, ..., m$$

- r is the number of DMUs, in our model equals 827
- s is the number of input, in our model equals 1
- m is the number of outputs, in our models equals 6

3.6.4 Result

Based on the analysis in R program after implementing our DEA model, we were able to obtain the ranking table of relative efficiencies for each university.

According to the ranking of our DEA model, the 1st university is A T Still University of Health Sciences and from the result, we can tell that most of the high ranking universities are among nursing and health-related industry. The reason behind this is because when building the model, the outcomes variables of high salary and median earnings are contributed significantly when determining the ranking and the average salary in the medical industry is high. As a result, in order to improve our model, more outputs of universities will be added to the outcomes, such as, number of published papers and number of Nobel Prize which based on the universesâĂŹ research abilities to give a more reasonable result.

Table 9: Final Rank by the Simple DEA Method

INSTNM	RelativeEfficiency
A T Still University of Health Sciences	1.000000
Albany Medical College	0.200000
Assabet Valley Regional Technical School	0.166667
Abington Memorial Hospital Dixon School of Nursing	0.142857
Adelphi University	0.111111
Adult and Community Education-Hudson	0.111111
Augsburg College	0.111111
Fordham University	0.111111
Mayville State University	0.111111
Holy Family University	0.090909

4 Optimization

After we calclate the score of of each college, then we need to optimize the ultimate effect by maximizing the toal efficiency of these investment. In this case, we will use the following equation to determin the amount of money and the duration of the investment of each school:

$$amount_i = \frac{score_i \times duration_i}{\sum_{i=1}^{N} score_i \times duration_i}$$

where $i = 1, 2...N$

We have used AHP to calculate the scores, and now we will introduce how we obtain the duation number.

4.1 Methodology

From the graph below, we can see that most of the shcools have donation less than $50,000. Thus, we will separate the data into different data sets according to its donation amount and then fit a polynomial model by using Multiple Linear Regression Model. When we choose best fit, we try the model from simple linear regression model to the multiple linear regression model with the fourth term. Then we determin the best model according to its R-Adjusted value and the significance of each term. We tend to pick the model with higher highest term because it will fit our data better. However, if the highest term is not siginifcant, we will not choose that model. We use this methodology in the model selection step for the following data sets analysis.

4.2 Data Set Separation

First, the range from 0 to 50000 is selected. Obvious patterns can be seen in the plot. Thus, we can separate the data points according to the density and moving trends shown in the plot. Here, $Investment = 7000$ and $Investmen = 27000$ are selected as the cutoffs points. In the second plot, we can see that most investment values are less than $150,000. Therefore, we will treat the points greater than $150,000 as outliers in this case.

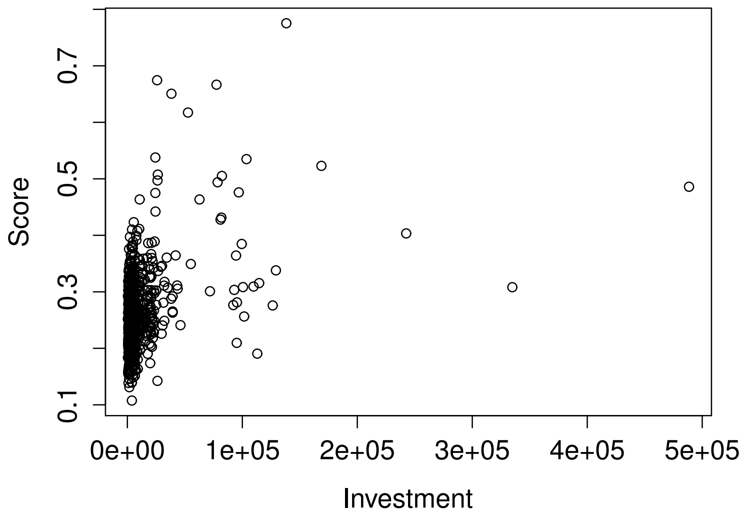

Figure 2: Score VS Investment

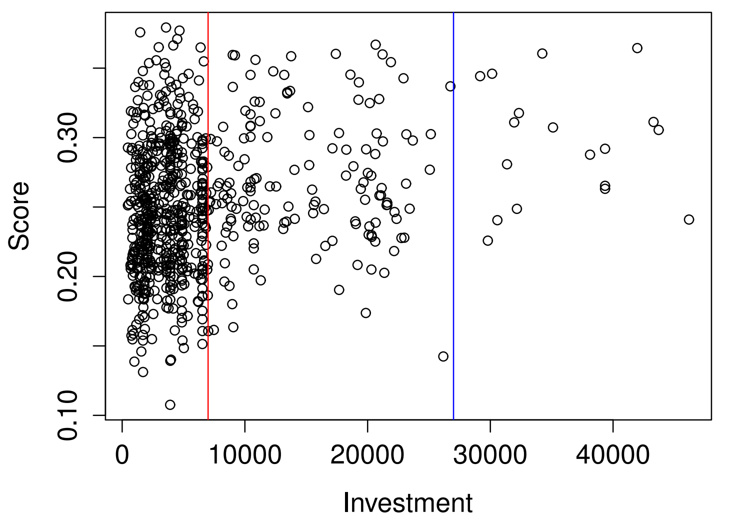

Figure 3: Score VS Investment Less than $50,000

4.2.1 Data Set 1

This data set contains all the points with investment less than $7,000, which means that these points represent those colleges which do not need much budget or which are not good enough for the foundations to invest in.

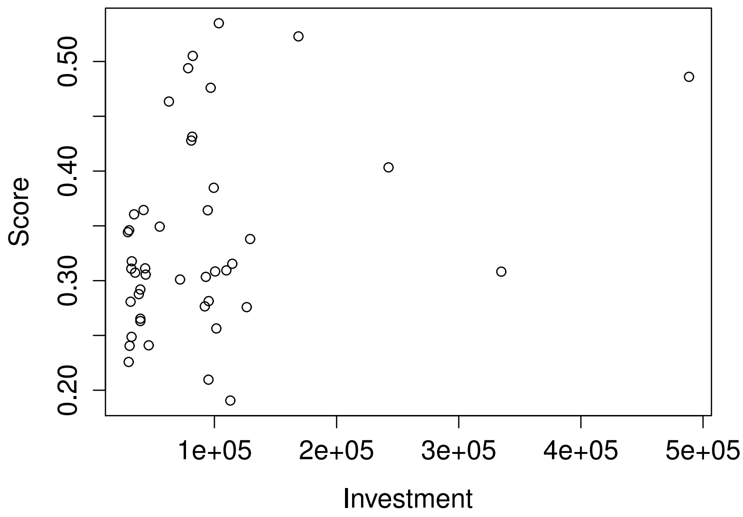

Figure 4: Score VS Investment between \$50,000 and \$500,000

The qudratic equation we obtained from R is the following:

$$y = -2.246 \times 10^{-9}x^2 + 1.1912 \times 10^{-5}x + .2163$$

where y is the calculated score and x is the investment amount. Here, we set

$$y' = 0$$

then the score reaches the maximum value when x=4256.46.

4.2.2 Data Set 2

This data set contains all the points with investment between \$7,000 and \$27,000, which means that these points represent those colleges which might be developing and needs the donation to a certain extent.

The cubic equation we obtained from R is the following:

$$y = 4.927 \times 10^{-14}x^3 - 2.681 \times 10^{-9}x^2 + 4.573 \times 10^{-5}x + 3.151 \times 10^{-2}$$

where y is the calculated score and x is the investment amount. Here, we set

$$y' = 0$$

and combining the plot, we will choose a point around 13000. After the calculation, we obtain that the score reaches the maximum value when x=14250.33.

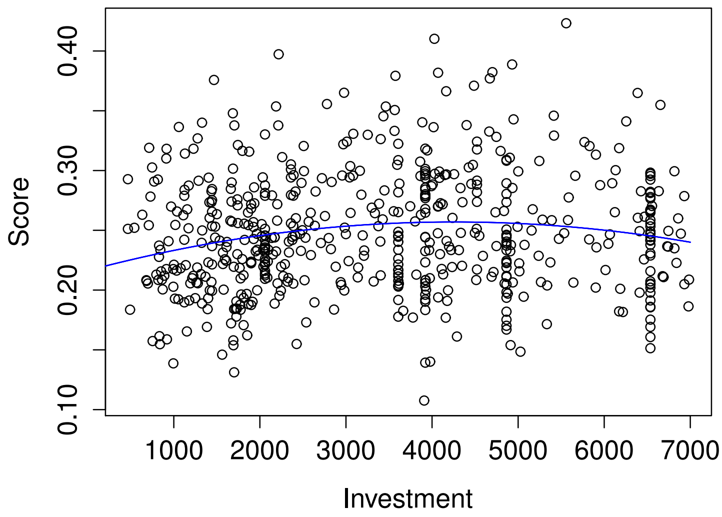

Figure 5: Fitted Quadratic Regression Line for Investment Less than $7,000

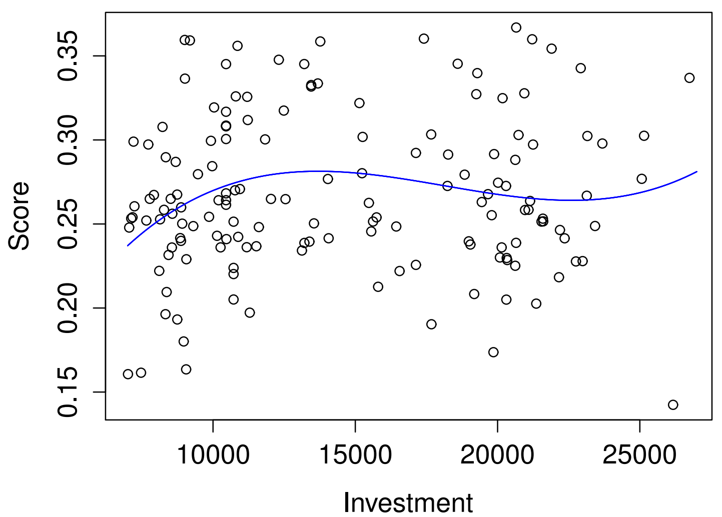

Figure 6: Fitted Cubic Regression Line for Investment Between $7,000 and $27,000

4.2.3 Data Set 3

This data set contains all the points with investment greater than $27,000 and less than $150,000, which means that these colleges needs a lot of donation to develop and it is very likely that some of them are medical schools or health centers. As we mentioned earlier, only four schools have donation more than $150,000, so we delete those points to fit a better regression for the rest of points.

The quartic equation we obtained from R is the following:

$$y = 4.555 \times 10^{-20}x^4 - 1.387 \times 10^{-14}x^3 + 1.427 \times 10^{-9}x^2 - 5.668 \times 10^{-5}x + 1.051$$

where y is the calculated score and x is the investment amount. Here, we set

$$y' = 0$$

and combining the plot, we will choose a point around \$75,000. After the calculation, we obtain that the score reaches the maximum value when x=75310.21.

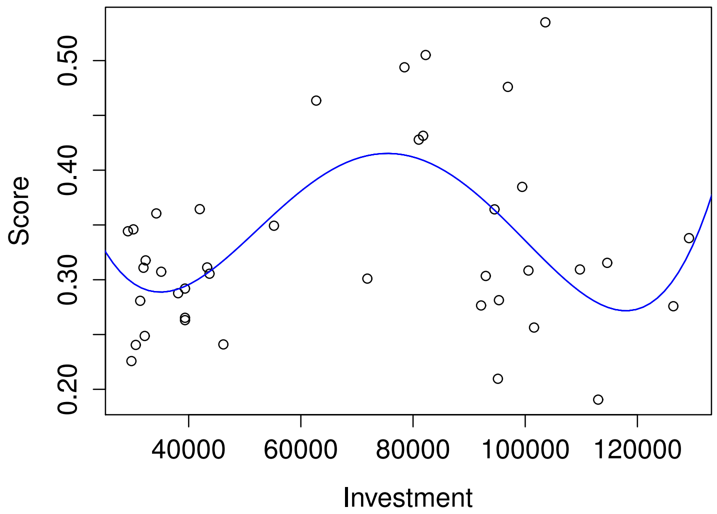

Figure 7: Fitted Quartic Regression Line for Investment Between \$27,000 and \$150,000

4.3 Result

Now, we update table and will have the following result:

Data Set Number	Range	Duration Factor
1	[0,7000)	4256.46
2	[7000,27000)	14250.33
3	[27000,150000)	75310.21

By using the equation we mentioned eariler,

$$amount_i = \frac{score_i \times duration_i}{\sum_{i=1}^{N} score_i \times duration_i}$$

where $i = 1, 2...N$

we obtain the following table (sorted by effiency):

Table 10: Summary of the Investment Strategy

INSTNM	Amount($M)	Duration	Explanaiton
Research College of Nursing	13.35	5	$2.67 million per year
Augustana College	14.07	5	$ 2.814 million per year
Rose-Hulman Institute of Technology	7.44	5	$1.488 million per year
University of Connecticut-Avery Point	1.38	1.4	About $1 million per year
Springfield College-School of Human Services	2.08	2.1	About $1 million per year
University of Connecticut-Tri-Campus	2.03	2.1	About $1 million per year
Gustavus Adolphus College	2.76	2.8	About $1 million per year
Akron School of Practical Nursing	1.24	1.3	About $1 million per year
University of Southern California	1.81	1.9	About $1 million per year
Strayer University-Arkansas	1.94	2	About $1 million per year
Brooklyn Law School	2.20	2.3	About $1 million per year
Texas Tech University Health Sciences Center	1.59	1.6	About $1 million per year
Cornell University	1.22	1.3	About $1 million per year
Susquehanna University	1.24	1.3	About $1 million per year
Chapman University	1.09	1.1	About $1 million per year
Mount Carmel College of Nursing	1.40	1.5	About $1 million per year
The University of Texas Health Science Center at Houston	1.45	1.5	About $1 million per year
Aquinas College	1.03	1.1	About $1 million per year
Luther Seminary	1.04	1.1	About $1 million per year
Colgate University	1.03	1.1	About $1 million per year
Tufts University	1.22	1.3	About $1 million per year
The University of Texas Health Science Center at San Antonio	1.47	1.5	About $1 million per year
Samford University	1.08	1.1	About $1 million per year
Louisiana State University Health Sciences Center-Shreveport	1.58	1.6	About $1 million per year
Lewis University	1.06	1.1	About $1 million per year
Michigan Technological University	1.06	1.1	About $1 million per year
Pepperdine University	1.34	1.4	About $1 million per year
Georgetown University	1.24	1.3	About $1 million per year
Washington and Lee University	1.50	1.6	About $1 million per year
Galen College of Nursing-Cincinnati	0.98	1	About $1 million per year
Augustana College	1.18	1.2	About $1 million per year
The Sage Colleges	1.12	1.2	About $1 million per year
Western University of Health Sciences	1.33	1.4	About $1 million per year
University of Georgia	0.97	1	About $1 million per year
Mount Mercy University	1.12	1.2	About $1 million per year
Creighton University	1.19	1.2	About $1 million per year
University of Phoenix-Kansas City Campus	1.04	1.1	About $1 million per year
Barnard College	1.10	1.1	About $1 million per year
Concordia College at Moorhead	1.08	1.1	About $1 million per year
Icahn School of Medicine at Mount Sinai	1.39	1.4	About $1 million per year
University of Jamestown	1.34	1.4	About $1 million per year
Carroll University	0.97	1	About $1 million per year
Samaritan Hospital School of Nursing	0.98	1	About $1 million per year
Thomas More College	1.14	1.2	About $1 million per year
Loma Linda University	1.41	1.5	About $1 million per year
Worcester Polytechnic Institute	1.19	1.2	About $1 million per year
Strayer University-Delaware	1.12	1.2	About $1 million per year
Claremont McKenna College	1.58	1.6	About $1 million per year
University of Florida	1.24	1.3	About $1 million per year
Massachusetts Maritime Academy	1.47	1.5	About $1 million per year
Dickinson State University	0.92	1	About $1 million per year
Central Susquehanna Intermediate Unit LPN Career	1.20	1.2	About $1 million per year

5 Conclusions

Our team implements three models to generate a new evaluation system and optimal investment strategy for Goodgrant Foundation.

First we use a simple triple-metric model to give a preliminary ranking of the colleges. Next, we modify the metric model based on AHP (Analytic Hierarchy Process) and TOPSIS (the Technique for Order of Preferency by Similarity to Ideal Solution). Then, we build another model based data envelopment analysis (DEA) to calculate relative efficiency of each school. Finally, we implement the multiple linear regression method to determine each schoolâĂŹs ranking score and the duration factor.

In the end, we discuss the strengths and weaknesses and write the proposal letter to the CEO of Goodgrant Foundation, Mr. Alpha Chiang about the final optimal investment strategy suggests that further variables based on the performance of faculties and schoolâĂŹs research abilities need to be taken into the consideration for further improvement.

6 Strengths and weaknesses

6.1 Strengths

- We use three different models to prove that our strategy generate similar results, and, thus, it is feasible.

6.2 Weaknesses

- The missing values can be predicted using other methods. However, we simply replace them by the average numbers. Thus, the result might not be accurate enough. It is important to ensure that the data is accurate.
- Take interest rate into consideration, and the results and the strategies can be imporved.

7 Letter to Mr. Alpha Chiang

FROM: Team 52063

TO: Mr. Alpha Chiang, Goodgrant Foundation

Dear Mr. Alpha Chiang,

In order to achieve the maximum return-on-investment (ROI) and the optimal investment strategy, our group divided the problem into two parts, 1) implementing the AHP and TOPSIS based multiple-metric model to rank the performance of universities by calculating the relative efficient usage of resources; 2) optimizing donation amount and time duration for TOP 51 universities by Multiple-Linear Regression.

Our group believes that the result of investment strategy is reliable because we firstly used a simple triple-matric model which arbitrarily picked top three important variable and then we built a DEA model to approach the problem with a completely different manner - the importance weights of each variable is calculated each time within the model to reduce the bias of each variable. Even though the approaches are different, the result is surprisingly similar, A T Still University of Health Sciences ranked No.1 in both ranking system.

Please refer to our attachment in the next page of this letter.

However, the concept of return-on-investment (ROI) for philanthropic educational investments requires further improvement in the future. For our model, we used the calculated efficiencies divided by total donation as the measurement of ROI. But the problem left to debate is, what are the necessary input variables included in the model to have a comprehensive strategy?

It is reasonable to come to the conclusion that the income of students after graduation places an important part when measure studentâĂŹs performance. As a result, the earnings of graduates are consistent with the final strategy; most high-rank universities are law schools, medicals school and nursing schools and it is not surprising that the graduates from those schools have the highest salaries. However, if the time is allowed, our group will include more variables as the measurement of universitiesâĂŹ research abilities since salaries are not the only standard for school performance. Variables we will consider further are faculty compensation and faculty with a PhD degree.

In conclusion, our group suggests that, even though the current models can generate the investment strategy based on the optimal ROI, more analysis based on the demand of universities can be included in the future philanthropic educational investments within the United States as well.

Sincerely,

Team 52063

References

[1] Alireza Afshari, M. M. (2010). Simple Additive Weighting approach to Personnel Selection problem. International Jouranl of Inovation, Management and Technology .

[2] Ata Allah Taleizadeh, S. T.-B. (2009). A hybrid method of Pareto, TOPSIS and genetic algorithm to optimize multi-product multi-constraint inventory control systems with random fuzzy replenishments. Mathematical and Computer Modeling , 1044-1057.

[3] Chen, T.-Y. (2012). Comparative analysis of SAW and TOPSIS based on interval-valued fuzzy sets: Discussions on score functions and weight constraints. Expert Systems with Application , 1848-1861.

[4] Olson, D. (2004). Comparison of Weights in TOPSIS Models. Mathematical adn Computer Modeling , 721-727.

[5] Zhi-Hua Hu, Jing-Xian Zhou, Meng-Jun Zhang and Yang Zhao. Methods for ranking college sports coaches based on data envelopment analysis and PageRank. Expert Systems, Volum 32, Issue 6, pages 652-673, December 2015.

[6] DEAzone.com. Retrieved January 2016, from `http://http://deazone.com/en/resources/tutorial`

[7] Lamport, Leslie, LaTeX: " A Document Preparation System ", Addison-Wesley Publishing Company, 1986.

Appendices

Appendix A R Code for Data Cleaing

```
data=read.csv(file="~/Desktop/CleanData.csv",header=TRUE,na.strings=c(".","#DIV/0!"))

allmean=array(dim=26)
for(i in 1:26){
  allmean[i]=mean(data[,i+3],na.rm=TRUE)
}

type=as.factor(data[,3])
mean=aggregate(data[,4:29],list(data$Type),mean,na.rm=TRUE)

for(i in 2:27){
  for(j in 1:29){
    if(is.na(mean[j,i])) mean[j,i]=allmean[i-1]
  }
}

data=data.frame(data)
length=length(data[,1])

for(i in 4:29){
  for(j in 1:length){
    if(is.na(data[j,i])){
      for(k in 1:29){
        if(mean$Group.1[k]==data[j,3]) data[j,i]=mean[k,i-2]
      }
    }
  }
}

write.csv(data,file="VeryGoodData.csv")
```

Appendix B R Code for DEA Model

```
library(lpSolve)

rm(list=ls())
data=read.csv("ExcellentData.csv",na.strings = ".")
inputs=data.frame(data[2])
outputs=data.frame(data[c(3,4,5,6,7)])
N=dim(data)[1] #number of DMUs
s=dim(inputs)[2] #number of input variable s= 1
m=dim(outputs)[2] #number of outputs m=4

for (i in 1:N)
{
  f.rhs <- c(rep(0,1,N),1)
  f.dir <- c(rep("<=",1,N),"=")
  aux <- cbind(-1*inputs,outputs)

  f.obj <- c(0*rep(1,s),as.numeric(outputs[i,]))
  f.con <- rbind(aux ,c(as.numeric(inputs[i,]), rep(0,1,m)))
```

```
  results <- lp ("max",as.numeric(f.obj), f.con, f.dir, f.rhs,scale=0, compute.sens=TRUE)
  if(i==1)
  {
    weights <- results$solution
    effcrs <- results$objval
    lambdas <- results$duals[seq(1,N)]
  }
  else
  {
    weights <- rbind(weights, results$solution)
    effcrs <- rbind(effcrs , results$objval)
    lambdas <- rbind(lambdas, results$duals[seq(1,N)] )
  }
}
```

Appendix C R Code for Optimization

```
data=read.csv(file="~/Desktop/Optimization.csv",header=TRUE,na.strings=c(".","#DIV/0!"))
x=data[,4]
y=data[,6]
plot(x,y,xlab="Investment",ylab="Score")

subdata1=data[which(data[,4]<7000),]
subdata2=data[which(data[,4]<50000  & data[,6]<0.38),]
plot(subdata2[,4],subdata2[,6],xlab="Investment",ylab="Score")
abline(v=7000,col=2)
abline(v=27000,col=4)

subdata3=data[which(data[,4]<27000 & data[,4]>=7000 & data[,6]<0.38),]
subdata4=data[which(data[,4]>=27000 & data[,4]<150000 & data[,6]<0.55),]
subdata4=data[which(data[,4]>=27000 & data[,6]<0.55),]
plot(subdata4[,4],subdata4[,6],xlab="Investment",ylab="Score")

plot(subdata1[,4],subdata1[,6])
plot(subdata2[,4],subdata2[,6])
plot(subdata3[,4],subdata3[,6])
plot(subdata4[,4],subdata4[,6])

remove('sx','sy')
sx=subdata4[,4]
sy=subdata4[,6]

lsx=sx
lsy=sy
lsx2=lsx^2
lsx3=lsx^3
lsx4=lsx^4

fourmod=lm(lsy~lsx+lsx2+lsx3+lsx4)
summary(fourmod)
curve(fourmod$coef[5]*x^4+fourmod$coef[4]*x^3+fourmod$coef[3]*x^2+fourmod$coef[2]*x+fourmod$coe

cubmod=lm(lsy~lsx+lsx2+lsx3)
summary(cubmod)
curve(cubmod$coef[4]*x^3+cubmod$coef[3]*x^2+cubmod$coef[2]*x+cubmod$coef[1],from=7000,to=27000,

quamod=lm(lsy~lsx+lsx2)
```

```
summary(quamod)
curve(quamod$coef[3]*x^2+quamod$coef[2]*x+quamod$coef[1],from=0,to=7000,add=T,col=4)

linmod=lm(lsy~lsx)
summary(linmod)
curve(linmod$coef[2]*x+linmod$coef[1],from=0,to=300000,add=T)
```