A Polychoric Correlation to Identify the Principle Component in Classifying Single Tuition Fee Capabilities on the Students Socio-Economic Database

by Wiyli Yustanti

Submission date: 15-Apr-2022 11:36AM (UTC+0700)
Submission ID: 1811203695
File name: Yustanti_2018_IOP_Conf._Ser.__Mater._Sci._Eng._288_012150.pdf (1,019.98K)
Word count: 3731
Character count: 18745

A Polychoric Correlation to Identify the Principle Component in Classifying Single Tuition Fee Capabilities on the Students Socio-Economic Database

W Yustanti* and Y Anistyasari

Informatics Engineering Department, Universitas Negeri Surabaya, Surabaya

*wiyliyustanti@unesa.ac.id

Abstract. The government has issued the regulation number 55 of 2013 about the enactment of a single tuition fee based on the socio-economic conditions of each student. All public universities are required to implement this policy. Therefore, each university needs to create a formulation that can be used to categorize a student into which cost group. The results of the data collection found that the parameters used to determine the classification of tuition fees between one universities with another are different. In this research, taken a sampling of student data at one public university which is using 43 predictor variables and 8 categories of single tuition. The sample data used are socioeconomic data of students of 2016 and 2017 classes received through public university entrance selections. The results of this study reveal that from 43 variables, there are 16 variables which are the most significant in influencing single tuition category with goodness-of-fit index is 0.866. This value means that the proposed model can indicate student's ability to pay the tuition fee.

1. Introduction

The consideration of the student's socio-economic status (SES) as a parameter in determining their tuition fee payment ability has been formalized by Indonesian government in the policy set forth in the Ministry of Education and Culture No. 55 Year 2013 [1]. In this rule is given classification of tuition starting from category 1 to 5 in accordance with the proposal of each college. To get a prediction of the student's ability to pay the tuition fees, the college provides a form with various attributes to measure SES. Several studies have shown that the number of parameters already in use to classify the tuition payment capability ranges from 3 to 14 attributes [2,3]. In this research will be used student data of 2016 and 2017 that entered into a state university through public university entrance selections [4,5]. In the registration process, students are asked to fill out a form that includes 43 variables used to see how their socio-economic status [6]. The main thing that will be analyzed in this research is what factors have a strong influence in determining socio-economic status [7,8]. The SES theory approach by using confirmatory factor analysis in previous research has been done has not yet produced a unique structural equation [9,10].

2. Tuition fee policy in Indonesia

The latest ministerial regulation relating to the single tuition fee (STF) policy was issued in May 2017 namely Ministerial Regulation no. 55 years 2017 on a single tuition at the state university [11]. STF is defined as the cost of each student based on his economic ability. STF consists of several groups that

are determined based on the economic capacity of students, parents, students or others who finance it [12,13,14]. Each state university proposes a STF grouping model to the finance minister in order to be formally established. The regulation also states that university leaders can provide STF assistance and / or reinstate STF to students if any [15,16]:

- Differences in the economic capacity of students proposed by students, parents, or others who finance them; And / or
- Changes in economic ability of students, parents, or others who finance it.

In addition, it is also stipulated that public universities are prohibited from collecting base money and /or other charges other than STF from new students of diploma and undergraduate programs for the benefit of direct learning services [17,18]. The university does not cover student fees consisting of personal fees, community service program fees, dorm fees and independent study and research activities.

3. Socio-Economic Status (SES)

Research on how to measure the validity and reliability of socioeconomic status has been widely practiced, especially research in the field of education, social and psychology. In relation to this study, the definition of the SES taken is a definition derived from the consortium of social scientists as follow [19,20].:

"SES can be defined broadly as one's access to financial, social, cultural, and human capital resources. Traditionally a student's SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighbourhood, and school resources"

The indicators based on the definition used in this study can be described in the table 1 below.

Table 1. Variable of Dataset as SES Indicator

Latent Variable		Indicator Variable		Measurement Scale
ξ_1	Family Income	x_1	Mother's employement	Ordinal
		x_2	Father's employement	Ordinal
		x_3	Mother's salary	Ordinal
		x_4	Father's salary	Ordinal
		x_5	Mother's other income	Ordinal
		x_6	Father's other income	Ordinal
ξ_2	House Value	x_8	House tenure	Ordinal
		x_9	Electricity Power	Ordinal
		x_{10}	Land Size	Ordinal
		x_{11}	House Size	Ordinal
		x_{12}	Landhouse Tax Value	Ordinal
		x_{13}	Roof Material	Nominal
		x_{14}	Floor Material	Nominal
		x_{15}	Wall Material	Nominal
		x_{16}	Wall Condition	Ordinal
		x_{17}	Livingroom Condition	Ordinal
		x_{18}	Roof Condition	Ordinal
		x_{19}	Bathroom Condition	Ordinal
		x_{20}	Kithcen Condition	Ordinal
		x_{21}	Guestroom Condition	Ordinal
		x_{22}	Family room Condition	Ordinal
		x_{23}	Bedroom Condition	Ordinal

Table 1. Variable of Dataset as SES Indicator

Latent Variable		Indicator Variable		Measurement Scale
		x_{24}	Balcony Condition	Ordinal
		x_{25}	Has Bathroom	Nominal
		x_{26}	Has Washing Area	Nominal
		x_{27}	Has Toilet	Nominal
ξ_3	Expenditure	x_{28}	Water Bill	Ordinal
		x_{29}	Electricity Bill	Ordinal
		x_{30}	Phone Bill	Ordinal
		x_{31}	Internet Bill	Ordinal
		x_{23}	Number of People at Home	Scale
		x_{33}	Motor Tenure	Ordinal
		x_{34}	Car Tenure	Ordinal
		x_{35}	Chilren are Schooling	Scale
		x_{7}	Number of Dependet	Ordinal
ξ_4	Parent's Education	x_{36}	Mother's education	Ordinal
		x_{37}	Father's education	Ordinal
ξ_5	Social and Culture	x_{38}	Is Father Alive	Nominal
		x_{39}	Father's Relationship	Nominal
		x_{40}	Is Mother Alive	Nominal
		x_{41}	Distance from City	Ordinal
		x_{42}	Source of Water	Nominal
		x_{43}	Source of Electricity	Nominal

4. Methods

Generally, the method used in this study can be explained in several stages as follows:

- Data Collections
- Measurement Modelling
- Model Validation
- Model Evaluation

4.1. Data collection

The data used in this experiment is taken from an academic database involving table schemes related to the new student registration process. In general, the field of the table used as many as 43 fields to be used as the indicator variable (independent) and 1 field as the dependent variable is the tuition fee level field. The total data used is 6059 records from student data in range 2016 -2017.

4.2. Measurement modelling

The variables consisting of observable or measured variables are called the manifest variables and the indirectly measured variables are called latent variables. Latent variables cannot be measured on a regular basis directly but can be established and built by other variables that can be measured. The variables used to construct latent variables are called indicator variables. In this study, there are 5 latent variables and 43 indicator variables as shown on table 1. The path constructed in each variable can be illustrated in figure 1 as initially structural relationship model. This model then evaluated by calculated the correlation and covariance value. The computation is conducted until reach the identified model which some parameter value successfully delivered [21,22].

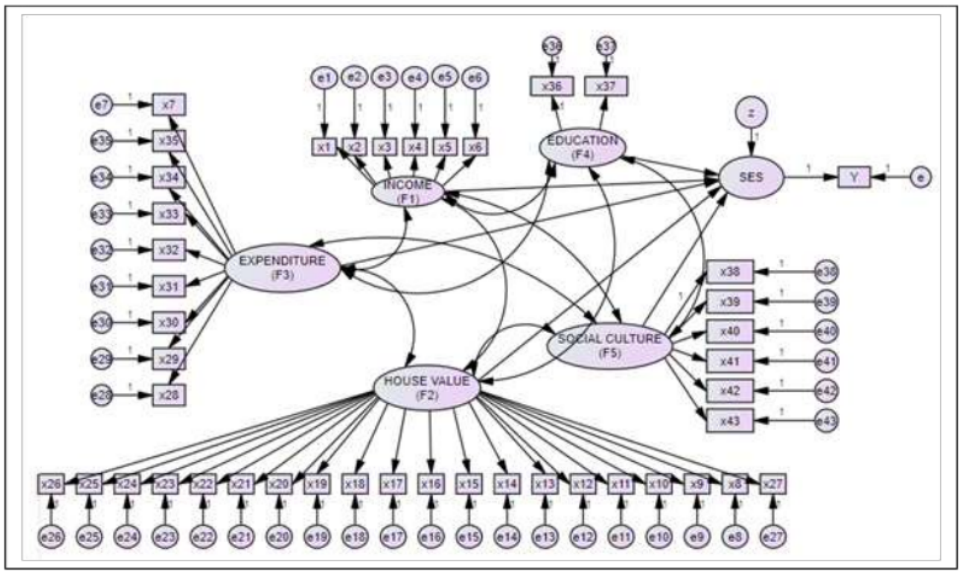

Figure 1. Initially Conceptual Model in Relationship Between SES Variables (model 1)

4.3. Model validation

The hypothesized model must be valid that refers to the ability of an indicator to measure what it really wants to be measured. The validity of an indicator is thus a requirement that must be met. The validity and indicators in measuring the latent variables were assessed by testing whether all their loading factor (λ_{ti}) were significant by using the t-test for a given level of α

4.4. Model evaluation

The first step in interpreting the resulting model is to assess whether the model is feasible or not. There is no single measure to judge the feasibility of a model. Suggests using the least three model feasibility tests. Here are some common model conformity sizes used to assess the feasibility of a model [23]:

- Chi-Square Test (χ^2), if the test value is significant compare to specified α
- GFI (Goodness of Fit Index). The general rule suggested for the feasibility of a model is its GFI value is greater than 0.90 and its maximum value is 1
- AGFI (Adjusted Goodness of Fit Index). A model is said to be good if its AGFI value is greater from 0.80 and its maximum value is 1
- RMSEA (Root Mean Square of Error Approximation), when RMSEA $\leq$ 0.08 then in general model already representing the actual data.

5. Results and discussion

In this research conducted many simulations in making conceptual model of data. Model 1 in Figure 1 is the first concept that involves all attributes contained in the database that supposedly affect the socioeconomic status of students, which in this case is indicated by the category of STF. The procedure generally begins by calculating the polychoric correlation of the data which will then be used to construct the covariance matrix as input from the confirmatory factor analysis. The confirmatory factor equation can be described as equation 1.

$$x = \Lambda_x \xi + \delta \quad (1)$$

where

x : vector of *q x 1* indicator variables
Λ_X : matrix for the loading factor (λ) or the coefficient which indicated relationship between x with ξ sized *q x n*
ξ : vector of *n x 1* latent variables
δ : vector of *q x 1* error measurement

The calculation of polychoric correlation and covariance of the data, conclude that there are 16 significant indicators that influence the student's SES.

5.1. Polychoric correlation matrix

The data used in this study is ordinal data, then the polychoric correlation matrix in the estimation of model parameters is the most appropriate matrix input. Using R-Language with polychor function in data processing, the table 2 is the correlation value based on the significant indicators.

Table 2. Polychoric Correlation of 16 Significant Indicator Variables

	X02	X04	X16	X17	X18	X19	X20	X21	X22	X23	X24	X29	X30	X31	X36	X37	Y
X02	1,00																
X04	0,58	1,00															
X16	0,18	0,29	1,00														
X17	0,16	0,30	0,84	1,00													
X18	0,19	0,31	0,79	0,72	1,00												
X19	0,20	0,35	0,73	0,72	0,72	1,00											
X20	0,23	0,38	0,75	0,74	0,72	0,85	1,00										
X21	0,17	0,28	0,85	0,84	0,75	0,75	0,77	1,00									
X22	0,19	0,31	0,83	0,82	0,76	0,77	0,80	0,90	1,00								
X23	0,19	0,31	0,80	0,78	0,78	0,78	0,80	0,83	0,86	1,00							
X24	0,17	0,29	0,78	0,79	0,73	0,73	0,75	0,83	0,82	0,80	1,00						
X29	0,24	0,48	0,24	0,28	0,23	0,33	0,36	0,23	0,26	0,27	0,26	1,00					
X30	0,28	0,47	0,24	0,24	0,25	0,30	0,32	0,24	0,26	0,26	0,26	0,46	1,00				
X31	0,28	0,48	0,24	0,27	0,24	0,30	0,32	0,24	0,25	0,27	0,25	0,48	0,72	1,00			
X36	0,26	0,37	0,15	0,18	0,15	0,23	0,25	0,15	0,18	0,17	0,16	0,39	0,38	0,41	1,00		
X37	0,37	0,54	0,17	0,19	0,18	0,25	0,27	0,16	0,20	0,20	0,18	0,44	0,40	0,42	0,69	1,00	
Y	0,51	0,63	0,31	0,35	0,32	0,40	0,44	0,32	0,35	0,36	0,33	0,71	0,50	0,54	0,49	0,50	1,00

5.2. Covariance matrix

If equation 1 is represented in matrix form, then it can be described as matrix in equation 2. The equation model of each measurement variable, for example x_2, then $x_2 = \lambda_{21}\xi_1 + \delta_2$. The covariance value also can be drawn by graphic in figure 2.

$$\begin{bmatrix} x_2 \\ x_4 \\ x_{16} \\ x_{17} \\ x_{18} \\ x_{19} \\ x_{20} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{24} \\ x_{29} \\ x_{30} \\ x_{31} \\ x_{36} \\ x_{37} \end{bmatrix} = \begin{bmatrix} \lambda_{21} & 0 & 0 & 0 \\ \lambda_{41} & 0 & 0 & 0 \\ 0 & \lambda_{162} & 0 & 0 \\ 0 & \lambda_{172} & 0 & 0 \\ 0 & \lambda_{182} & 0 & 0 \\ 0 & \lambda_{192} & 0 & 0 \\ 0 & \lambda_{202} & 0 & 0 \\ 0 & \lambda_{212} & 0 & 0 \\ 0 & \lambda_{222} & 0 & 0 \\ 0 & \lambda_{232} & 0 & 0 \\ 0 & \lambda_{242} & 0 & 0 \\ 0 & 0 & \lambda_{293} & 0 \\ 0 & 0 & \lambda_{303} & 0 \\ 0 & 0 & 0 & \lambda_{314} \\ 0 & 0 & 0 & \lambda_{364} \\ 0 & 0 & 0 & \lambda_{373} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} + \begin{bmatrix} \delta_2 \\ \delta_4 \\ \delta_{16} \\ \delta_{17} \\ \delta_{18} \\ \delta_{19} \\ \delta_{20} \\ \delta_{21} \\ \delta_{22} \\ \delta_{23} \\ \delta_{24} \\ \delta_{29} \\ \delta_{30} \\ \delta_{31} \\ \delta_{36} \\ \delta_{37} \end{bmatrix} \quad (2)$$

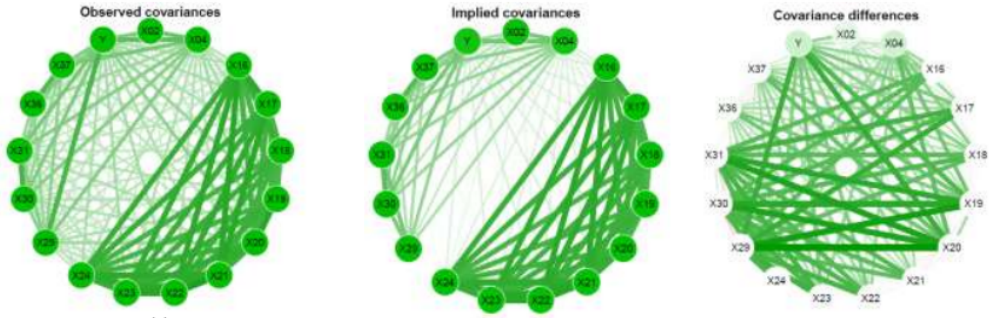

Figure 2. (a) Covariance Value (b) Implied Covariance (c) Covariance Differences a -b

5.3. Enhancement measurement model

The improvement model for the path diagram for the strong variables in determining the SES is the income (ξ_1), house value (ξ_2), and parent's education (ξ_3) and expenditure (ξ_4) variables. The path analysis diagram completely can be illustrated in figure 3. The loading factor value a can be seen on the path coefficient. Commonly, the λ value which is more than 0,6 is considered as influential variables. The computation results to identify the structural model is displayed in table 3.

Table 3. Parameter Estimates in Model 2

	Estimate	Std Error	Z Value	Pr(>|z|)	Path
λ_{2}	0.60382977	0.012631740	47.802581	0.000000	x02 <--- INCOME
λ_{4}	0.89994852	0.012420831	72.454778	0.000000	x04 <--- INCOME
λ_{16}	0.89908109			0.000000	x16 <--- HOUSEVALUE
λ_{17}	0.88361504	0.010065152	87.789534	0.000000	x17 <--- HOUSEVALUE
λ_{18}	0.83153698	0.010445788	79.605000	0.000000	x18 <--- HOUSEVALUE
λ_{19}	0.83757634	0.010403509	80.509019	0.000000	x19 <--- HOUSEVALUE
λ_{20}	0.86046308	0.010238918	84.038475	0.000000	x20 <--- HOUSEVALUE
λ_{21}	0.92901644	0.009702144	95.753723	0.000000	x21 <--- HOUSEVALUE
λ_{22}	0.93361097	0.009663739	96.609704	0.000000	x22 <--- HOUSEVALUE
λ_{23}	0.90507031	0.009897341	91.445808	0.000000	x23 <--- HOUSEVALUE
λ_{24}	0.87752514	0.010111585	86.784131	0.000000	x24 <--- HOUSEVALUE
λ_{29}	0.66475988	0.012234391	54.335346	0.000000	x29 <--- EXPENDITURE
λ_{30}	0.79793626	0.011555023	69.055358	0.000000	x30 <--- EXPENDITURE
λ_{31}	0.82351582	0.011431018	72.042210	0.000000	x31 <--- EXPENDITURE
λ_{36}	0.77806952	0.012420831	64.733940	0.000000	x36 <--- EDUCATION
λ_{37}	0.86602688	0.012019499	73.145517	0.000000	x37 <--- EDUCATION

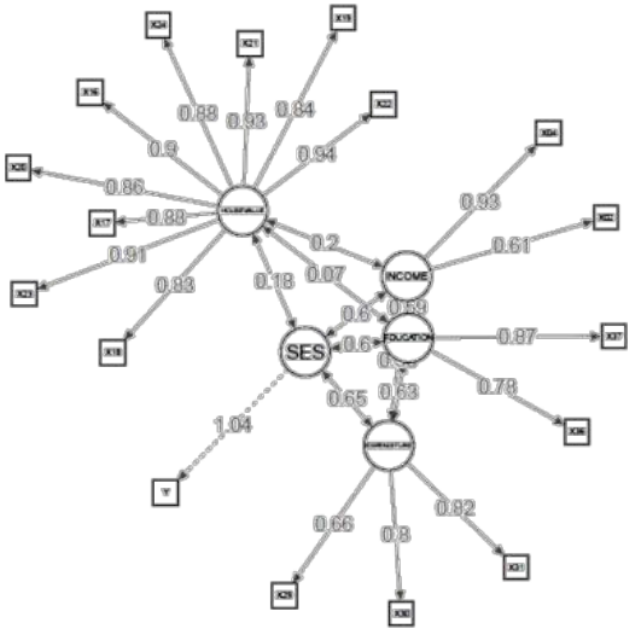

Figure 3. Path diagram for model 2 (enhancement model)

5.4. Goodness of fit

From the second modelling results can then be seen how the value of parameters for the conformity test model of the data. The value of Chi-Square test obtained for the model in Fig. 3 is 7948,795, with

degrees of freedom being 112, means that the model does not represent well the relationships contained in the sample, or it can be said that the model is not consistent with the relationship that occurs in the actual data. The value of GFI obtained is 0.866, this value is close to 1, meaning that the proposed model is fairly good. While the value of AGFI obtained is 0.8172, this value is also close to 1, meaning that the proposed model is good enough. When seen from the value of RMSEA obtained is 0.10 then it means that the proposed model is fairly good. Other parameters that can be used are Bentler-Bonett NFI = 0.9159956, Tucker-Lewis NNFI = 0.8992872, Bentler CFI = 0.91706, Bentler RNI = 0.91706 and Bollen IFI = 0.9170811, where each is close to 1. Thus, overall, the model obtained in the results of this study is sufficient to represent the data.

6. Conclusion

The results of this study have been able to develop an SES model as an indication of the ability to pay the tuition. However, generally the resulting model can still be upgraded to close to 90%. Another approach in reducing the dimensions of variables that contribute to the prediction of dependent variables can also be performed. The result of this study is further investigated in next step which is the classification purposes with variety methods of data mining.

References

[1] Ilham T 2016 Model Pengambilan Keputusan Penentuan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri *Journal Speed – Sentra Penelitian Engineering dan Edukasi* **8** 2

[2] Rizky R C and Setiawan E 2013 Ketakbiasan Dalam Model Analisis Faktor Konfirmatori (CFA) Pada Metode Pendugaan Kuadrat Terkecil Terboboti (Weighted Least Square) untuk Data Ordinal *Prosiding Semirata* **1** 1

[3] Huseynpur, Babak, Masoud Y M and Ghafour R 2015 The Relationship among Variables of Students" Socio-economic Status" *International Journal of Educational Investigations* **2** 1 80-92

[4] Roberts J K 1999 *Basic Concepts of Confirmatory Factor Analysis Reporting Speech* Texas

[5] Bradley R H and Corwyn R F 2002 Socioeconomic status and child development *Annual review of psychology* **53** 1 371-399

[6] Budianto K, Steven R, Sentinuwo, Alwin M S 2017 Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan Data Mining *E-Journal Teknik Informatika* **11**

[7] Betts J and Roemer J E 2005 Equalizing opportunity for racial and socioeconomic groups in the United States through educational finance reform *Department of Economics, UCSD*

[8] Ensminger M E, Forrest C B, Riley A W, Kang M, Green B F, Starfield B and Ryan S A 2000 The validity of measures of socioeconomic status of adolescents *Journal of Adolescent Research* **15** 3 392-419

[9] White K R 1982 The relation between socioeconomic status and academic achievement *Psychological bulletin* **91** 3 461

[10] Dwi N P and Rostika L 2017 Kaji Banding Metode Topsis, SAW dan AHP-Topsis guna Menentukan UKT Mahasiswa Baru Di Politeknik Negeri Cilacap *Jurnal Infotekmesin* **8**

[11] Regulation of the minister of research, technology and high education of the republic of Indonesia number 39 year 2017 about the single tuition fee for public universities Jakarta

[12] Ariyady K M and Made S 2015 Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggal Mahasiswa Baru *Jurnal Lontar Komputer* **6**

[13] Muhammad J 2016 *Sistem Pendukung Keputusan untuk Penentuan Kelompok Uang Kuliah Tunggan dengan Metode K Mean Clustering* Malang, Undergraduate Thesis

[14] Yustanti W and Anistyasari Y 2017 Latent Variable Reconstruction in Determining Student's Single Tution Fee Category with Confirmatory Factor Analysis Approach submitted on *The 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*

[15] Syaiful R, Imam F R and Rosa A A 2017 Pengembangan Sistem Penunjang Keputusan Penentuan UKT Mahasiswa dengan menggunakan Metode Moora Studi Kasus Politeknik Negeri Malang *Jurnal Informatika Polinema* **3** 4
[16] Etis S, Siska Y and Syahrul A 2015 Analisis Tingkat Uang Kuliah Tunggal dengan Menggunakan Regresi Logistik Ordinal *Jurnal Gradien* **11** 2 1096-1100
[17] Ulumudin M I 2006 *Penggunaan korelasi polikhorik pada data berskala ordinal dalam analisis model persamaan struktural* (Studi Kasus: Analisis Kepuasan Terhadap Kualitas Layanan Cyber Mahasiswa IPB)
[18] Sirin S R 2005 Socioeconomic status and academic achievement: A meta-analytic review of research *Review of educational research* **75** 3 417-453
[19] U S Department of Education 2012 Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress:A Theoritical Foundation , Recommendations to the National Center for Education Statistics, Institute of Education Sciences, National Center for Education Statistics
[20] Carnevale A P and Rose S J 2003 Socioeconomic Status, Race/Ethnicity and Selective College Admissions *A Century Foundation Paper*
[21] Gusmelia T and Rachmansyah 2017 Pemanfaatan Metode Simple Additive Weighting (SAW) untuk Penentuan Penerima UKT Kelompok 1 *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)* 9 2579-5406
[22] Muhammad F 2016 *Pengembangan Sistem Penentuan Uang Kuliah Tunggan dengan Metode Fuzzy C Means* Malang, Undergraduate Thesis
[23] Kohei A 2016 *Matrix-Based Introduction to Multivariate Data Analysis* Graduate School of Human Sciences, Osaka, Japan, Springer

A Polychoric Correlation to Identify the Principle Component in Classifying Single Tuition Fee Capabilities on the Students Socio-Economic Database

ORIGINALITY REPORT

20%	18%	9%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.science.gov Internet Source	7%
2	Submitted to Universitas Pendidikan Indonesia Student Paper	5%
3	Submitted to Universitas Sebelas Maret Student Paper	3%
4	files.eric.ed.gov Internet Source	2%
5	nces.ed.gov Internet Source	2%
6	cyberleninka.org Internet Source	2%

Exclude quotes	Off	Exclude matches	< 2%
Exclude bibliography	On		