

# The Implementation of Nonlinear Principal Component Analysis to Acquire the Demography of Latent Variable Data (A Study Case on Brawijaya University Students)

Solimun<sup>\*</sup>, Adji Achmad Rinaldo Fernandes, Retno Ayu Cahyoningtyas

Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Malang, Indonesia

Received March 22, 2020; Revised April 24, 2020; Accepted May 3, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Nonlinear principal component analysis is used for data that has a mixed scale. This study uses a formative measurement model by combining metric and nonmetric data scales. The variable used in this study is the demographic variable. This study aims to obtain the principal component of the latent demographic variable and to identify the strongest indicators of demographic formers with mixed scales using samples of students of Brawijaya University based on predetermined indicators. The data used in this study are primary data with research instruments in the form of questionnaires distributed to research respondents, which are active students of Brawijaya University Malang. The used method is nonlinear principal component analysis. There are nine indicators specified in this study, namely gender, regional origin, father's occupation, mother's occupation, type of place of residence, father's last education, mother's last education, parents' income per month, and students' allowance per month. The result of this study shows that the latent demographic variable with samples of a student at Brawijaya University can be obtained by calculating its component scores. The nine indicators formed in  $PC_1$  or  $X_1$  were able to store diversity or information by 19.49%, while the other 80.51% of diversity or other information was not saved in this  $PC$ . From these indicators, the strongest indicator in forming latent demographic variables with samples of a student of Brawijaya University is the origin of the region ( $I_2$ ) and type of residence ( $I_5$ ).

**Keywords** Nonlinear Principal Component Analysis, Mixed Scales, Demographic

## 1. Introduction

Multivariate analysis is one type of statistical analysis

that simultaneously analyzes multiple variables on an individual or object [1]. By using the multivariate analysis, the influence of several variables toward other variables can be done simultaneously for each object of research [5]. Based on the measurement process, variables can be categorized into manifest variables (observable) and latent variables (unobservable) [2][3]. Generally, latent variables are defined by the variables that cannot be measured directly, yet those variables must be through the indicator that reflects and constructs it [11]. Latent variables can be filtered into factual latent variables, such as demographic variables that will be examined in this study [4]. Demography is research regarding the human population in a particular area, especially about the composition of a community and its development [10]. The indicator model that forms or constructs the variables is known as the formative indicator model wherein this indicator model that constructs it is not obliged to have common factors [7]. [10] To measure the data of latent variables using the formative indicator model, it can be used primary component score acquired through the Principal Component Analysis.

## 2. Materials and Methods

### 2.1. Research Materials

The measurement result can be categorized into two groups of data, which are quantitative and qualitative data [8]. The quantitative data is an observation result measured within a numerical scale while the qualitative is an observation result in a form of a category instead of numbers. [9] Based on its measurement scales, data can be categorized into four types, including nominal, ordinal, interval, and ratio. The nominal and ordinal scales are

included in the category of qualitative data (non-metric). The internal and ratio scales are included in the category of quantitative data (metric)[2], as follows:

1. Nominal data comprises of naming elements. The word of nominal derived from Latin *nomos* which means name. To use nominal data, it is required scores of “1” for men and “0” for women an alternative into numerical form by scoring, thus, it can be analyzed using statistics. An example of this case is that the change of sex category comprises of men and women by giving the scores of “1” for men and “0” for women.
2. Ordinal data is the data that does not only contain naming elements but also order elements. The word ordinal derived from the word ‘order’. For instance, for the data of education that comprises elementary, middle, and high schools, constructed order and level are from the lowest to the highest.
3. Interval data is the data that does not only have naming and order elements but also interval elements. For example, there was research with students as the subject, some variables that can be categorized into interval data are the Grade Point Average (GPA) and height.
4. Ratio data is an interval data wherein the interval has a meaning (the zero is absolute). An example of this data category is the income variable [2].

## 2.2. Research Method

The data used in this research is the primary data from active undergraduate students at Brawijaya University collected by questionnaires. The examined latent variables being demography with used nine indicators. The nine indicators comprise of five nominal data-scaled indicators (Sex X1 (men, women); Place of origin X2 (Malang Raya, East Java, Outside of Java); Father’s occupation X3.

The variables in this research include groups of indicators and latent variables. The indicator groups comprise of Sex ( $I_1$ ), Place of Origin ( $I_2$ ), Father’s Occupation ( $I_3$ ), Mother’s Occupation ( $I_4$ ), Type of Housing ( $I_5$ ), Father’s Latest Education ( $I_6$ ), Mother’s Latest Education ( $I_7$ ), Parents’ Total Monthly Income ( $I_8$ ), and Students’ Monthly Pocket Money ( $I_9$ ). Meanwhile, the study latent variables are demographic variables.

This research functions to collecting the latent variables data using a mix-scaled indicator with latent variables by acknowledging the demographic characteristics of the chosen indicator being. There are used nine indicators in this research with different scales, which respectively have multiple categories as well. The scales and categories of each indicator that have been determined will be explained in the following Figure 1.

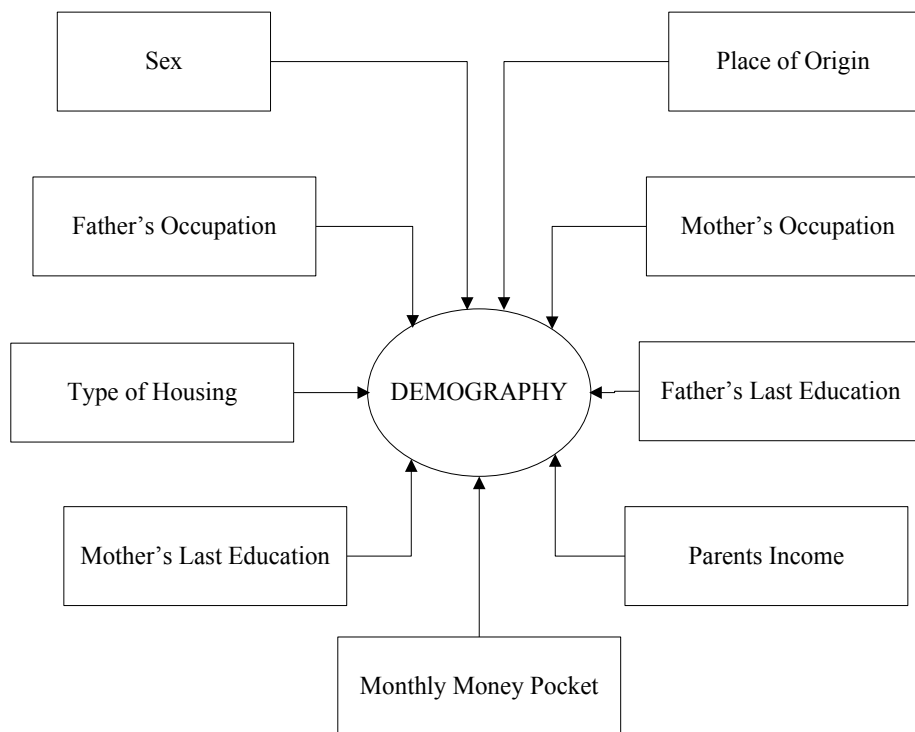


Figure 1. Indicators and Latent Variables

**2.3. Research Steps**

The following is an analysis step carried out in this research:

1. Determining research variables, indicators, and designing research instruments.
2. Collecting data from respondents using the research instrument.
3. Determining score of category quantification.
4. Determining eigenvalue and component weight.
5. Determining the equation model of the first principal component.
6. Acquiring data of the transformation result.
7. Determining data of demographic latent variables by calculating the component score.
8. Concluding.

**3. Results and Discussion**

The Nonlinear Principal Component Analysis used in this research has a mixed scale that comprises of nominal, ordinal, and ratio measurement scales. The used indicator and data scale can be seen in Table 1.

**Table 1.** Indicator Measurement Scale

Indicator	Measurement Scale
Sex ( $I_1$ )	Nominal
Place of Origin ( $I_2$ )	Nominal
Father's Occupation ( $I_3$ )	Nominal
Mother's Occupation ( $I_4$ )	Nominal
Type of Housing ( $I_5$ )	Nominal
Father's Latest Education ( $I_6$ )	Ordinal
Mother's Latest Education ( $I_7$ )	Ordinal
Parents' Total Monthly Income ( $I_8$ )	Ordinal
Monthly Money Pocket ( $I_9$ )	Ratio

Based on Table 1, It can be acknowledged that Sex ( $I_1$ ), Place of Origin ( $I_2$ ), Father's Occupation ( $I_3$ ), Mother's Occupation ( $I_4$ ), and Type of Housing ( $I_5$ ) have a nominal scale because they only have naming elements. Meanwhile, Father's Latest Education ( $I_6$ ), Mother's Latest Education ( $I_7$ ), Parents' Total Monthly Income ( $I_8$ ), and an ordinal measurement scale because they have both named and order elements. Monthly Pocket Money ( $I_9$ ) has a ratio scale because it has an order, distance, and value of absolute zero.

**Table 2.** Frequency Distribution of Sex Indicator

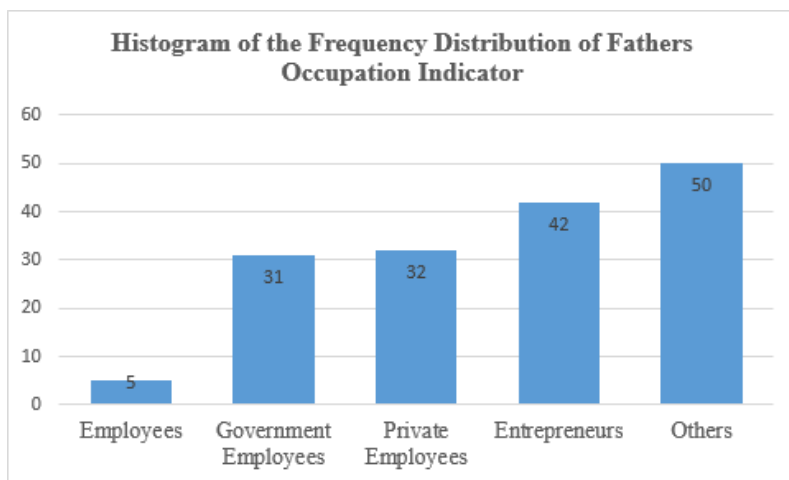
Sex	Frequency
Men	69
Women	91
<b>Total</b>	<b>160</b>

Based on Table 2, it can be acknowledged that students who become the sample in this research are 160 wherein 91 of the respondents are women and the other 69 were men. Therefore, it can be concluded that female respondents are more dominant.

**Table 3.** Frequency Distribution of Place of Origin

Place of Origin	Frequency
Malang Raya	35
East Java (other than Malang Raya)	58
Java (Other than East Java)	45
Outside of Java	22
<b>Total</b>	<b>160</b>

According to Table 3, the respondents mostly came from the area of East Java (other than Malang Raya) with a total number of 58 students, followed by the respondents from Jawa Tengah, Jawa Barat, and Banten as many as 45 students, the respondents from Malang Raya are 35 students, and the respondents from Outside of Java are 22 students.



**Figure 2.** Frequency Distribution of Father's Occupation Indicator

The histogram in Figure 2. shows that 5 respondents have fathers who work as employees, 31 respondents have fathers who work as government employees, 32 respondents have fathers who work as private employees, 42 respondents have fathers who work as entrepreneurs and the other 50 have fathers who have other occupations besides the ones mentioned before. The research finding regarding the mother's occupation indicator is presented in Figure 3.

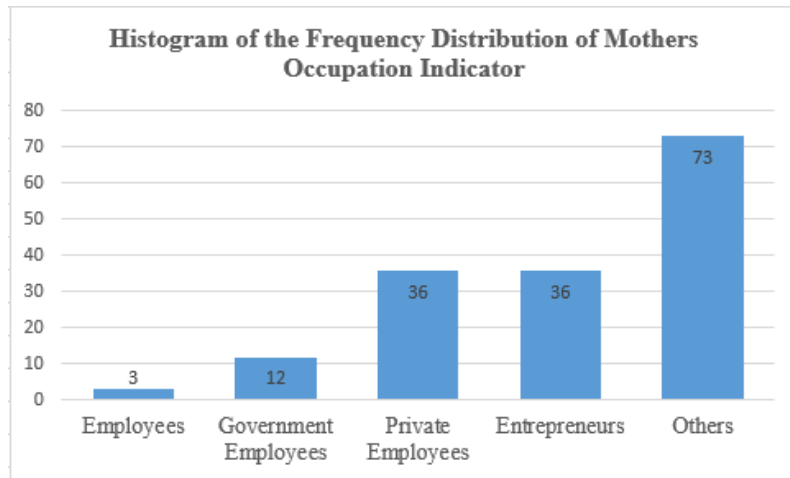


Figure 3. Frequency Distribution of Mother's Occupation Indicator

The histogram in Figure 3. shows that 36 respondents have mothers who work as private employees and 36 respondents have mothers who work as entrepreneurs. Furthermore, 12 respondents have mothers working as employees and the other 73 respondents have mothers who have other occupations besides the ones mentioned before.

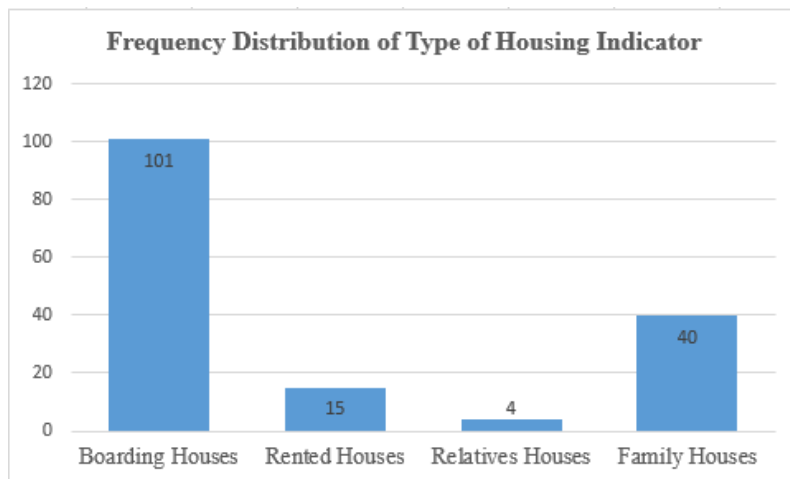


Figure 4. Histogram Frequency Distribution of Type of Housing Indicator

The histogram in Figure 4 above shows that 101 respondents are currently live in boarding houses, 15 respondents are currently residing in rented houses, 4 respondents are currently residing in houses owned by relatives (other than parents and biological siblings), and the other 40 respondents are currently residing in family houses. Therefore, it can be acknowledged that the majority of the respondents in this research are students who are living in boarding houses.

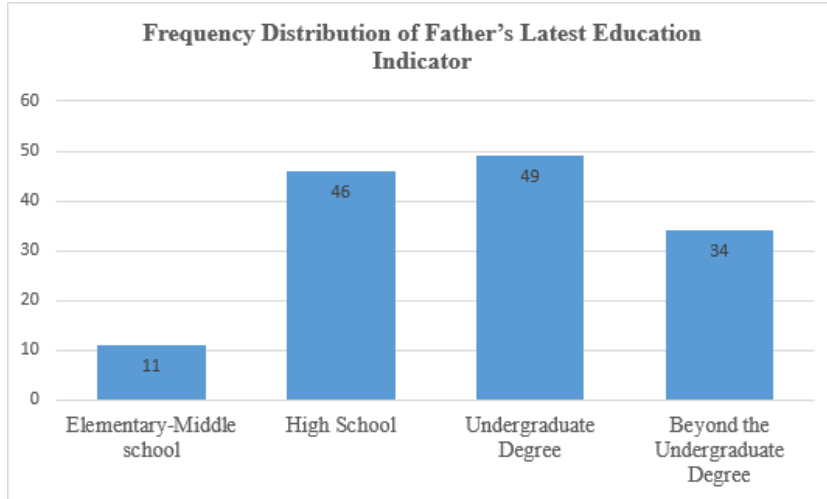


Figure 5. Histogram Frequency Distribution of Father's Latest Education Indicator

For Father's Education indicator, it can be acknowledged that 69 respondents have fathers whose latest education is High Education (Undergraduate Degree), 11 and 46 respondents have fathers whose latest education is, respectively, elementary-middle school and high school, as well as the rest 34 respondents have fathers whose latest education is beyond the undergraduate degree.

The descriptive analysis for the indicator of Mother's Education can be seen in Figure 6.

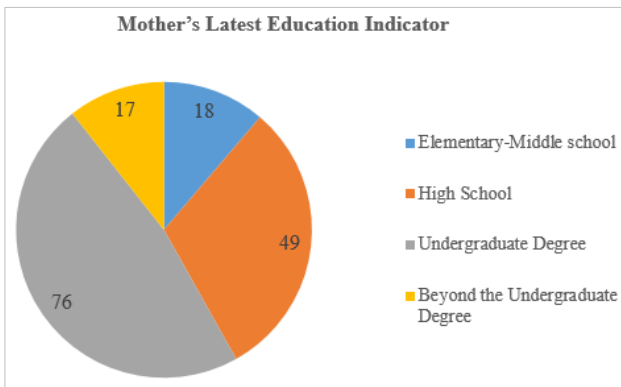


Figure 6. Pie Chart of Mother's Latest Education Indicator

Based on the graphic in Figure 6, it can be seen that 76 respondents have mothers whose latest education is High Education (Undergraduate Degree), 18 and 49 respondents have mothers whose latest education is, respectively, elementary-middle school and high school, as well as the rest 17 respondents have mothers whose latest education is beyond the undergraduate degree.

Figure 7 elaborates that 45 respondents have parents with an income under Rp5,000,000.00 monthly, then 50 respondents have parents who earn Rp5,000,000.00 up to Rp7,500,000.00 monthly, and the rest, which is 29 respondents, has parents whose income is Rp7,500,000.00-Rp. 10,000,000.00 and 36 respondents have parents with an income above Rp10,000,000.00 each

month.

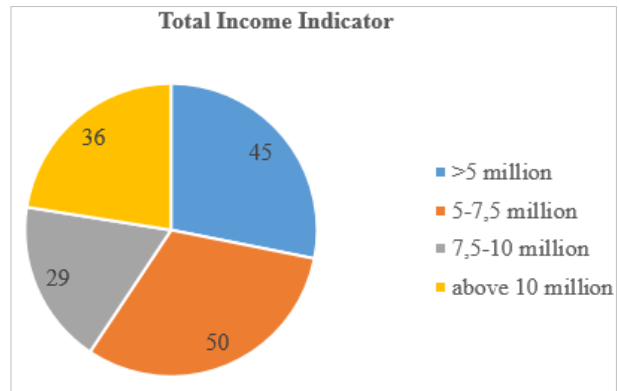


Figure 7. Pie Chart of Total Income Indicator

Table 4. Descriptive Analysis of Monthly Pocket Money Indicator

Descriptive Analysis	Value
Minimum	Rp0,00
Maximum	Rp4,000,000.00
Median	Rp1,225,000.00
Mean	Rp1,308,333.00
Deviation Standard	Rp689,888.00

Figure 8 and Table 4 show that the minimum pocket money achieved by the respondents in this research is Rp0,00 each month and the maximum pocket money received by the respondents are as much as Rp4,000,000.00 each month. The average monthly pocket money from 160 respondents in this research is Rp1,308,333.00 with a deviation standard of Rp686,888.00 wherein this pocket money has a median value of Rp1,225,000.00. it means that 80 student respondents received pocket money under Rp1,225,000.00 monthly and the rest of 80 students have pocket money more than that

nominal.

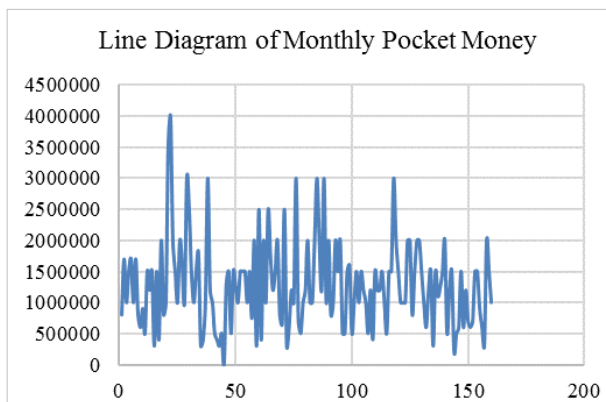


Figure 8. Line Diagram of Monthly Pocket Money Indicator

## 4. Conclusions

Based on the findings achieved, the conclusion is the data of demographic latent variables with mixed scales using students of Brawijaya University as the sample can be acquired by calculating the principal component score through the Nonlinear Principal Component Analysis. The nine indicators with the respective component weight formed in the  $PC1$  and constructing the demographic latent variables can keep the variance or information as much as 19.49%, while the other 80.51% of the variance or information is not stored in the  $PC1$ . The data of this principal component score can be used for further analysis, as the discriminant analysis, path analysis, cluster analysis, etc. The strong indicators in creating the demographic latent variables are the indicators of the place of origin ( $I_2$ ) and the type of housing ( $I_5$ ).

---

## REFERENCES

- [1] Ahsan, H. (2010). Analisis Gerombol Berhira pada Data Kategorik. Bogor: Institut Pertanian Bogor.
- [2] Fernandes, A.A.R. and Soehono, L. A. (2006). Kajian Analisis Cluster pada Data Berskala Campuran. Penelitian DPP. Malang. Universitas Brawijaya.
- [3] Fernandes, A.A.R., Hutahayan, B., Solimun, Arisoelaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, and Amaliana, L., (2019). Comparison of Curve Estimation of the Smoothing Spline Nonparametric Function Path Based on PLS and PWLS In Various Levels of Heteroscedasticity. *IOP Conference Series: Materials Science and Engineering*. Forthcoming Issue.
- [4] Fernandes, A.A.R., Darmanto, Solimun, Nurjannah, Amaliana, L., Astuti, A.B., Yanti, I., Arisoelaningsih, E., and Isaskar, R. (2019). Smoothing Spline Nonparametric Path: Application for Green Product and Green Marketing Strategy towards Green Product Purchasing Intention. *IOP*

*Conference Series: Earth and Environmental Science*. Vol 239. No 012018.

- [5] Fernandes, A.R., Solimun, Nurjannah, and Hutahayan, B. (2020). Comparison of use of linkage in integrated cluster with discriminant analysis approach. *International Journal of Advanced Science and Technology*. 29(3), pp. 5654-5668.
- [6] Gifi, A. (1981). *Nonlinear Multivariate Analysis*. Leiden: Universitas Leiden.
- [7] Hidayat, M.F., Fernandes, A.A.R., and Solimun, (2019), Estimation of Truncated Spline Function in Nonparametric Path Analysis Based on Weighted Least Square (WLS). *IOP Conference Series: Materials Science and Engineering*. Forthcoming Issue.
- [8] Islamiyati, A. D. (2010). Aplikasi Komponen Utama Nonlinier (PRINCALS) pada Peningkatan Mutu Pendidikan Tinggi. Makassar: Universitas Hasanuddin.
- [9] Johnson, R. A. and Wichern, D.W. (2002). *Applied Multivariate Analysis*. New Jersey: Prentice Hall Inc.
- [10] Mantra, I. B. (2000). *Demografi Umum*. Yogyakarta: Pustaka Pelajar.
- [11] Solimun. (2020). Warppls analysis application for effects of personality and commitment on the engagement of ngrebeg mekotek traditional actors in munggu village, bali. *International Journal of Advanced Science and Technology*. 29(4), pp. 2025-2044.