Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 20 No.2, Oktober 2015, p: 68-77

available online at:
journal.ipb.ac.id/index.php/statistika

NONLINEAR PRINCIPAL COMPONENT ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS WITH SUCCESSIVE INTERVAL IN K-MEANS CLUSTER ANALYSIS

Arista Marlince Tamonob[1], Asep Saefuddin[1], Aji Hamim Wigena[1]

[1]Department of Statistics, Bogor Agricultural University, Indonesia

ABSTRACT

K-Means Cluster is a cluster analysis for continuous variables with the concept of distance used is a euclidean distance where that distance is used as observation variables which are uncorrelated with each other. The case with the type data that is correlated categorical can be solved either by Nonlinear Principal Component Analysis or by making categorical data into numerical data by the method called successive interval and then used Principal Component Analysis. By comparing the ratio of the variance within cluster and between cluster in poverty data of East Nusa Tenggara Province in K-Means cluster obtained that Principal Component Analysis with Successive interval has a smaller variance ratio than Nonlinear Principal Component Analysis. Variables that take effect to the clusterformation are toilet, fuel,and job.

Keywords: K-Means Cluster Analysis, Nonlinear Principal Component Analysis, Principal Component Analysis, Successive interval.

INTRODUCTION

In the social and behavioral science , researchers are often confronted with a large number of variables. Often, variables used has a different measurement scale, there are numerical and categorical scale. Cluster analysis is a method in multivariate to grouping a set of objects in the same group (cluster) are more similar to each other than to those in other groups (clusters). In general, variables used as basic cluster consist of two types, there are categorical variables (nominal and ordinal) and numerical variables (interval and ratio).

In general, there are two methods in cluster analysis, hierarchical and non-hierarchical method. Hierarchical method is used if the number of cluster previously unknown while non hierarchical method is used if the number of cluster already known. One of non-hierarchical cluster analysis commonly used is the analysis of K-Means cluster used for numerical data and used euclid distance as a measure to see similarity and dissimilarity measured. Euclidean distance is used if the observation variables are independent oruncorrelated with one another.

One of the assumptions in cluster analysis is no multicoliniearity between variables (Hair, 2010). Euclid distance concept used for numerical data by transforming the variables using Principal Component Analysis (Kaufma & Petter 2005), while case with categorical data can be resolved by the develop method by Gifi in 1989 that is Nonlinear Principal Component Analysis. The outcome of Nonlinear Principal Component Analysis are score of component object and principal component that has numerical data which can be analyzed with cluster analysis using the concept of euclidean distance (Safitri 2012).

Nominal and ordinal data can be a interval by using the method proposed by Hays (1976) called successive interval. Principal Component Analysis can be used in the results of successive interval to reduce data and solve multicolliniearity, so it can be analyzed by cluster analysis.

Previous researches related to cluster analysis for categorical data was done by Ahzan (2010) which research about hierarchical cluster in categorical data used Nonlinear Principal Component Analysis and has result that complete linkage better than single and centroid linkage to explained the real condition. For mixed data, Islamiyati (2010) did research about application nonlinear principal component analysis to increase the quality of education in university used K-Means cluster , Yuniato

(2013) did research about the comparison between cluster analysis with nonlinear principal component analysis and two step cluster in mix data and the result that two step cluster method is better and more spesific than centroid linkage method with principal component analysis transformation. This study did research about the comparison between nonlinear principal component analysis and principal componet analysis with successive interval for categorical data in K-Means cluster analysis using poverty data in East Nusa Tenggara Province.

Categorical Data

Based on the type of the data , it can be divided into numerical data (quantitative) and categorical data (qualitative). Numerical data is data expressed in numerical magnitude (numbers), for example the data in per capita income, expending, price, etc, whereas categorical data is classified by category or a particular class.

Categorical data consist of nominal data and ordinal data. Nominal data is data with order or the value does not indicate the level merely as a label only, such as religion, gender, ethnicity or race. Ordinal data is data with the order of categories shows level or rank, such as education level, smoking habits. Categorical data can also be obtained by grouping continuous data but the risk is can lose information. In practice, it is easier to record categorical data rather than continuous data, respondents are easier to answer sensitive questions, it is more practical significance. Presentation of categorical data can be either frequency, frequency tables and contingency tables (Agresti 2002).

In cases where the numerical variables are linked, correlation analysis is one option to see the relation between variables. However, if the two categorical variables are linked, then correlation analysis can not be used because the numbers on a category code is not the only form of actual values. Another reason why the correlation analysis can not be used on categorical data because one type is a categorical variable that can not be sorted nominal category. Giving a different order will give different correlation value so that two people who calculate the correlation value likely give the same results. For this reason the chi-squared analysis that will be used to find relationships (associations) between the categorical variables. Analysis is based on the chi-square contingency table (often called cross-tabulations). Contingency table is a table whose cells contain the frequencies of the intersection of rows and columns. The general form of contingency tables with the first variable has *m* categories and the second variable has *k* categories. The hypotheses:

H_0: two variables are independent (associated)

H_1: two variables are dependent (not associated)

Statistical testused:

$$\chi^2_{hitung} = \sum_{j=1}^{k}\sum_{i=1}^{m}\frac{\left[O_{ij}-E_{ij}\right]^2}{E_{ij}}$$

If χ^2_{hitung} valueis greater than $\chi^2_{tabel(db;\alpha)}$ with the value ofthedegrees of freedom $(m-1)(k-1)$ then reject H_0.

Strong or weak associations can be seen from the association between the interval -1 to 1.Values association equals to 1 which means there is astrong association among variables, if the association equals to the value of zero, it means there is no association among variables.

Table 1 Correlation test or the choosen association

Variable 1	**Variable 2**	Correlation test or the choosen association
Nominal	Nominal	Contingency coefficient, Lambda, Phi, and Cramer
Nominal	Ordinal	Contingency coefficient, Lambda
Ordinal	Ordinal	Spearman, Gamma, Somers'd
Ordinal	Numeric	Spearman
Numeric	Numeric	Pearson

K-Means Cluster Analysis

K-Means cluster analysis has been developed by Mac Queen in 1967 and it is one of non-hierarchical data grouping methods that is most well-known and widely used in many fields because it is simple and easy to implement (Wichern & Johnson 2002). K-Means is a partitioning cluster method that separates data into different groups. The purpose of grouping the data is to minimize the variance within a group and maximize the variance between groups. Basic K-Means algorithm is as follows:

1. Determine the *k* value as the number of cluster to be formed
2. Generate the center point of the initial *k* randomly cluster
3. Calculate the distance of each data into each cluster center using Euclid distance
4. Group each data based on the distance between the nearest inter-data center
5. Determine new cluster center position by calculating the average value of the data existing on the same cluster center.

Principal Component Analysis

Principal component analysis is a multivariate method that aims to reduce the dimension of original variables to new variables (called principal component) that uncorrelated but have big information about original variables.

Principal component analysis is a weighted linear combination of origin variables are able to explain data maximizely.

The principal component *j* of *p* variables can be expressed as follows:

$$Y_j = a_{1j}x_1 + a_{2j}x_2 + ... + a_{pj}x_p = \underline{a}'\underline{x} \tag{2}$$

and the variance of principal components *j* is

$$Var(Y_j) = \lambda_j;\ j = 1, 2, ..., p \tag{3}$$

$\lambda_1, \lambda_2, ..., \lambda_p$ is eigen values that obtained from equation:

$$\left|\Sigma - \lambda_j \mathrm{I}\right| = 0 \tag{4}$$

Where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Eigen vector $\underline{a}$as the weight of transformation linear of origin variables obtained from the equation:

$$\left(\Sigma - \lambda_j \mathrm{I}\right)\underline{a}_j = 0 \tag{5}$$

Variance total ofprincipal component is

$$\lambda_1 + \lambda_2 + ... + \lambda_p = tr\left(\Sigma\right)$$

and the percentage of variance total data explained by the principal component *j* is:

$$\left(\frac{\lambda_j}{tr(\Sigma)} x100\%\right)$$

The percentage of total variance is considered sufficiently representative of variance if data is 75% or more.

Successive interval

Methods to convert data from nominal or ordinal data to interval data (Waryanto 2006).

Steps that must carried out are:

a.Calculate the frequency of each data category

b. Calculate the proportion based on the frequency of each category

c. Calculate the cumulative proportion

d. Calculate the value of for each cumulative proportion

e. Calculate the value of the density functionz

f. Calculate the value scale (the average interval) for each category

g. Calculates caling

Nonlinear Principal Component Analyis

Nonlinear Principal Component Analysis (NLPCA) is an extension of Principal Component Analys is called PRINCALS (Principal Component Analysis by Alternating Least Squares) or Principal Component Analysis using Alternating Least Square approach introduced by Gifi in 1989. NLPCA produces weight variable (variable loadings), quantification category (category quantifications) and a component score of the object (object scores). In NLPCA, the category of all variables with the scale not be given a numerical quantification of the category with an appropriate numerical scale. NLPCA aims to optimize or find the average quadratic optimal correlation between variables which have been given to the component quantification category. In search of the optimal value, both components of loading and quantification category will give various value until optimum value is found (Konig 2002 referred to Azizah et al 2013).

If there is a data that is formed into the matrix size to facilitate the calculation of NLPCA used notation:

n = the number of observations (object)
m = the number of variables
k_j = the number of categories j variables; $j = 1, 2, ..., m$
h_j = vector of j column of matrix H with size nx1
G_j = Indicator matrix of h_j with size nxk_j where:

$$g_{(j)ir} \begin{cases} 1, \text{ if object i is belongs to category r of j variables} \\ 0, \text{ if object i not belongs to category r of j variables} \end{cases}$$

$\underline{G}_j$ is a matrix that contains the frequency of each category at each variable. $\underline{G}_j$ complete when every line on one element having another one and is zero. $i = 1, 2, ..., n;\ r = 1, 2, ..., k_j$

Nonlinear Principal Component Analysis based on a meet loss theory that aims to minimize the homogeneity function loss (σ_M)

$$\sigma_M(X, Y_1, ..., Y_m) = m^{-1} \sum_{j=1}^{m} (\underline{X} - \underline{G}_j \underline{Y}_j)'(\underline{X} - \underline{G}_j \underline{Y}_j)$$ (8)

with normalization of the AVE $(X_s) = 0$ to the dimensions=1, .., p and X'X= I. AVE (X_s) is a vector which is the average ofthe columnsof the matrix elements X_s. The notation σ_M is used because the implications for the rank value of $\underline{G}_j$ least p.

σ_M function minimized using alternating least squares method to estimate the value of the weight variable a_j. The algorithm for calculating the first p dimensions simultaneously and minimizing using alternating least squares is:

$Y_j = D_j^{-1} G_j^{'} X$ with $D_j = G_j{}' G_j$

$$Z = \sum_{j=1}^{m} G_j Y_j$$

$$X = GRAM(Z)$$

X is a score of component object $n \times p$ (p = many variables), Y_j is multiple category quantification $k_j \times p$, GRAM (Z) is an orthogonal matrix of orthogonality Gram-Schmidt of the matrix Z, and a_j is weighting variable to variable j (Gifi 1989).

MATERIAL AND METHOD

Data

The data used is the core data SUSENAS 2013 of East Nusa Tenggara Province. Variables used consisted of 15 variables categorical: Education (X_1), Floor Area(X_2), Floor Type(X_3), Wall(X_4), Roof(X_5), Toilet(X_6), Water(X_7), Fuel(X_8), Assets(X_9), Literacy(X_{10}), Employment (X_{11}), Work Status(X_{12}), Credit(X_{13}), Poor Rice(X_{14}),Handphone(X_{15}).
Groups in this analysis consists of the poorandnon-poorhouseholds.

Method

In this research, steps to analyze are:

1. Data exploration
2. Checked multicolliniearity among variables
3. Applied NLPCA to data
4. Carried out the Successive interval to data
5. Applied PCA to successive interval data
6. Applied K-Means Cluster Analysis on principal component scores of NLPCA and PCA with Successive interval
7. Compared the two methods by ratio of variance within and between clusters

RESULTS

Data Description

Description SUSENAS core data of East Nusa Tenggara Province in 2013 to 10422 households.The following three variables were taken for description of the data:

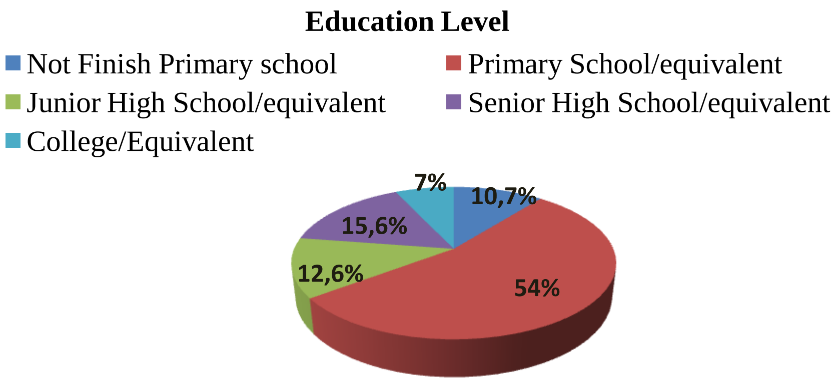

Figure 1 Percentage of level of education of household head

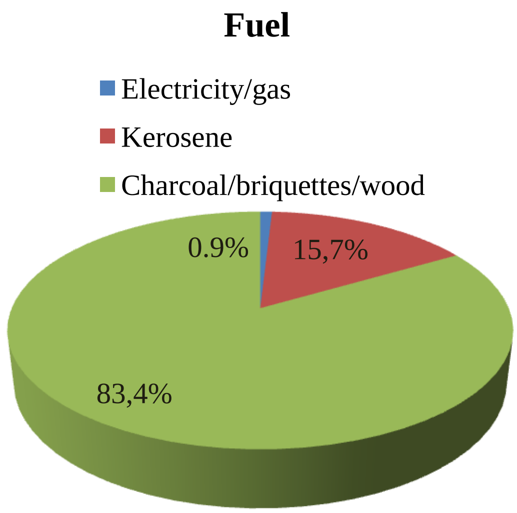

Figure 2 Percentage of the type of fuel used for cooking

From figure 1, we can see 54% of household heads in East Nusa Tenggara Province 2013 seek ededucation only up to primary level/equivalent and had 7% of college education/ equivalent.

From figure 2, we see that household fuel is still dominated by charcoal / briquettes / wood by 83.4%, electricity / gas by 0.9% and 15.7% kerosene.

We see in figure 3 that percentage of households with the wall of the house is made of walls by 32.7% and not made of walls by 32.7%.

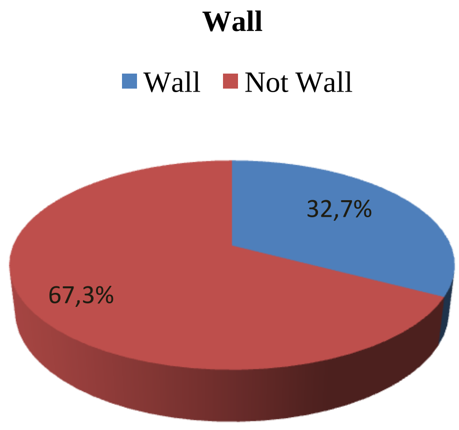

Figure 3 Percentage of walls of the house occupied by households

Checking Multicoliniearity

From the 15 independent variables of nominal and ordinal scale the test multicoliniearity assumptions was done by looking at the association between variables in the contingency tables.

Test of the association among the variables was large enough association between the variables of literacy and education by 0.815, between jobs and education by 0.656, between– the working status and education by 0.567, and between jobs and working status by 0.813.

Eigen values are shown in Table 1 in each ofthe principal components used to determine the number of major components used in the analysis. NLPCA results showed that when the two principal components were taken to mean, only 36% could explain the variance of the initial data. To be able to represent the variance of data in this study were taken 9 principal components with the total variance of 79%. The end result is the NLPCA was principal component scores were used for the analysis of K-Means cluster.

Nonlinear Principal Component Analysis Method

Table 1 Result of nonlinear principal component analysis

Principal Component	Eigen Value	% Variance	% Cummulative Variance
1	3.815	25.434	25.434
2	1.607	10.712	35.146
3	1.226	8.176	44.322
4	1.012	6.750	51.072
5	0.957	6.380	57.452
6	0.932	6.210	63.662
7	0.854	5.692	69.354
8	0.753	5.021	74.375
9	0.713	4.751	79.126
10	0.693	4.623	83.749
11	0.657	4.378	88.127
12	0.581	3.873	92
13	0.564	3.757	95.757
14	0.394	2.628	98.385
15	0.242	1.615	100

Table 2 Principal component score

	KU_1	KU_2	KU_3	KU_4	KU_5	KU_6	KU_7	KU_8	KU_9
X_1	-0.714	0.083	0.363	0.035	-0.106	0.228	0.156	-0.051	0.037
X_2	-0.363	-0.576	-0.391	-0.019	0.001	0.093	0.124	-0.034	-0.244
X_3	-0.435	-0.145	-0.221	-0.182	0.492	0.398	-0.043	-0.26	0.136
X_4	0.579	0.244	0.22	0.183	-0.173	-0.023	-0.065	-0.072	-0.086
X_5	0.539	0.383	0.057	-0.005	-0.041	0.156	0.327	-0.13	0.423
X_6	0.471	0.374	0.021	-0.084	0.308	0.144	-0.032	-0.487	-0.3
X_7	0.445	0.078	0.096	0.03	0.014	0.672	0.214	0.477	0.144
X_8	0.642	-0.259	-0.074	0.157	-0.149	0.019	-0.161	-0.063	-0.324
X_9	-0.329	0.404	0.388	-0.152	0.398	-0.027	-0.349	0.337	-0.214
X_{10}	0.416	0.207	0.596	0.139	0.291	0.299	0.06	0.259	0.194
X_{11}	-0.676	0.491	-0.287	0.226	-0.139	0.05	-0.051	-0.018	-0.006
X_{12}	-0.47	0.531	0.454	0.319	0.204	0.064	0.087	0.064	0.151
X_{13}	0.005	-0.174	0.13	0.802	0.489	-0.128	0.221	0.055	0.025
X_{14}	- 0.457	0.187	0.048	-0.258	0.048	-0.311	0.624	0.047	-0.285
X_{15}	0.599	0.183	-0.112	-0.084	0.096	-0.069	0.235	-0.019	-0.207

The principal component score is a score that indicates the amount of value or contribution of each principal component of each unit of observation. The principal component score can be positive or negative. Positive values mean a principal component has a big contributed and has positive influence on the observation unit and vice versa (Handoyo & Setiawan 2009). As an example of Table 2 above, for the variables X1 value of the first principal component is -0714 means that the principal component give a contribution 0.714 but has negative influence on the observation unit, a second component is 0.083 and has positive influence on the observation unit. This also applies to principal components and other variables.

Principal Component Analysis with Successive interval

Results of principal component analysis showed that when the two main components which were taken to mean, only 35.2% could explain the variance of the initial data. To be able to represent the variance of the data in this study were taken 9 main components with the total variance of 79.1%. The final results of principal component analysis is a score of major components used for the analysis of K-Means cluster.

Table 3 Result of principal component analysis of successive interval data

Principal Component	Eigen Value	Variance Proportion	Cummulative Variance Proportion
1	3.7140	0.248	24.8%
2	1.5729	0.105	35.2%
3	1.2984	0.087	43.9%
4	1.0467	0.070	50.9%
5	0.9811	0.065	57.4%
6	0.9432	0.062	63.6%
7	0.8481	0.057	69.3%
8	0.7543	0.050	74.3%
9	0.7160	0.048	79.1%
10	0.6941	0.046	83.7%
11	0.6606	0.044	88.1%
12	0.5827	0.039	92.0%
13	0.5545	0.037	95.7%
14	0.3616	0.024	98.1%
15	0.2808	0.019	100%

Table 4 Principal component score of successive interval data

	KU_1	KU_2	KU_3	KU_4	KU_5	KU_6	KU_7	KU_8	KU_9
X_1	0.372	-0.040	-0.378	-0.201	0.169	-0.121	0.134	-0.073	0.012
X_2	0.201	0.417	0.375	-0.063	0.069	-0.088	0.128	-0.076	-0.312
X_3	0.229	0.085	0.206	0.175	-0.154	-0.638	0.040	-0.288	0.237
X_4	-0.302	-0.174	-0.208	-0.214	0.037	0.132	-0.099	-0.065	-0.019
X_5	0.284	-0.287	-0.069	-0.082	0.104	-0.079	0.354	-0.099	0.519
X_6	-0.244	-0.296	-0.070	0.115	-0.150	-0.302	-0.007	-0.578	-0.336
X_7	-0.229	-0.065	0.076	-0.264	0.234	-0.563	0.244	0.578	-0.166
X_8	-0.328	0.213	0.058	-0.228	0.070	0.041	-0.184	-0.104	-0.349
X_9	0.163	-0.328	-0.319	0.304	-0.294	-0.207	-0.327	0.366	-0.291
X_{10}	0.236	0.159	-0.539	-0.330	0.263	-0.112	0.062	-0.238	-0.162
X_{11}	0.308	-0.426	0.277	-0.257	0.034	0.068	-0.070	-0.020	0.014
X_{12}	0.221	-0.430	0.358	-0.409	0.053	0.063	-0.137	-0.084	-0.137
X_{13}	-0.000	0.145	-0.088	-0.464	-0.830	0.023	0.239	0.066	0.032
X_{14}	0.235	-0.170	-0.022	0.292	-0.006	0.269	0.680	0.041	-0.304
X_{15}	-0.312	-0.136	0.081	0.061	-0.014	0.004	0.285	-0.093	-0.302

From Table 4 above for the variables X1 value of the first principal component is 0.372 means that the principal component give a contribution 0.372 but has positive influence on the observation unit, a second component is 0.040 and has negative influence on the observation unit. This also applies to principal components and other variables.

Compare Clustering Results

The best clustering results if object within cluster has low variance while object between cluster has high variance. From both method has different variance either variance between cluster and variance witin cluster.

Table 5 The comparison of cluster variance from each method

Variance	Method	
	NLPCA	PCA with successive interval
Within Cluster	3.895989	0.988311
Between Cluster	0.050125	0.020909
Rasio	77.72546	47.26725

From Table 5, it can be seen that the ratio between the variance in the cluster of both methods showed that the method of Principal Component Analysis method with Successive interval has ratio 47.26725 smaller than Nonlinear Principal Component Analysis the amounted to 77.2546. It explains that the method of Principal Component Analysis with Successive interval better than the method of Principal Component Analysis Nonlinear.

Table 6 ANOVA of K-Means cluster for principal component analysis with successive interval

Variables	Deviation standard within cluster	Deviation standard between cluster	F	p-value
X_1	0.2527	0.0320	0.06	0.8083
X_2	0.2125	0.0409	0.15	0.7134
X_3	0.3213	0.0395	0.06	0.8140
X_4	0.1547	0.0064	0.01	0.9364
X_5	0.2048	0.1125	1.21	0.3142
X_6	0.1628	0.2004	6.06	**0.0489**
X_7	0.1420	0.0478	0.45	0.5260
X_8	0.2129	0.3993	14.06	**0.0095**
X_9	0.3599	0.0745	0.17	0.6932
X_{10}	0.2658	0.2270	2.92	0.1385
X_{11}	0.2062	0.2312	5.03	**0.0662**
X_{12}	0.4184	0.0629	0.09	0.7739
X_{13}	0.2207	0.0489	0.20	0.6733
X_{14}	0.2351	0.2200	3.50	0.1104
X_{15}	0.1839	0.0398	0.19	0.6802

ANOVA results showed that the variables that differentiate in cluster formation is variable with p-value lessthan 0.1. In this case the significant variables are X_6(toilet), X_8(fuel), X_{11}(job).

CONCLUSION

In this study, it was concluded that NLPCA and PCA with successive interval produced 9 principal components with variance total about 79%. In K-Means cluster analysis with PCA successive interval had variance ratio within cluster and between cluster smaller than NLPCA about 47.26725. Variables whichhas influenced the household clustering were X_6 (toilet), X_8 (fuel), X_{11} (job).

From this research we can see that both of this method can used for categorical data but the different of this two methods is in NLPCA can't transform to the origin variables if we used the object scores while in PCA we can transform it back way and from the successive interval we can get the numerical data, so we can use in another analysis using numerical data.

REFERENCES

Abilawa M Sabeth. 2010. *Studi Determinan Karakteristik Rumah Tangga Miskin Provinsi Nanggroe Aceh Darussalam (Analisis Data SUSENAS)*. Jakarta: Universitas Indonesia.

Ahzan H. 2010. Analisis Gerombol Berhirarki pada Data Kategorik [skripsi]. Bogor(ID): Institut Pertanian Bogor.

Agresti A. 2002. *An Introduction to Categorical Data Analysis*. New York (US): John Wiley and Sons.

Azizah M. Soehono L. Solimun. 2013. *Analisis Cluster Komponen Utama Nonlinier dan Analisis Two Step Cluster untuk Data Berskala Campuran*. Malang: Universitas Brawijaya.

[BPS] Badan Pusat Statistik. 2014. *Nusa Tenggara Timur dalam Angka 2014*. Kupang: BPS Nusa Tenggara Timur.

De Leeuw J. Mair P. 2009. *Gifi Methods for Optimal Scaling in R: The Package Homals. Journal of Statistical Software*. 31(4). University of California. Los Angeles.

Dian. E. 2013. *Macam-Macam Kriteria Kemiskinan* [Internet]. [diunduh 2015 Mei 6]; Tersedia pada:http://eliadian.blogspot.com/2013/03/macam-macam-kriteria-kemiskinan.html.

Gifi. A. 1989. *Nonlinear Multivariate Analysis*. Chichester: John Wiley & Sons.

Hair J F. Black W C. Babin B J. Anderson R E. 2010. *Multivariate Data Analysis. Seventh Edition*. New Jersey: Prentice Hall International Inc.

Handoyo E, Setiawan A. 2009. Analisis Kebutuhan Perangkat Lunak menggynakan Analisis Faktor pada Program Studi Ilmu Keperawatan UNDIP. *Teknik*. 30(1):30-38.

Hays W L. 1976. Quantification in Psychology. Prentice Hall: New Delhi.

Johnson R. Wichern D. 2002. *Applied Multivariate Statistical Analysis.* Prentice Hall: New Jersey

Kaufma L. Peter J R. 2005. Finding Groups in Data. An Introduction to Cluster Analysis. New York: John Wiley & Sons Inc.

Kuroda M. Mori Y. Masaya I. Sakakihara M. 2011. *Alternating least squares in nonlinear principal components*. The Japan Society for the Promotion of Science (JSPS). Grant-in-Aid for ScientificResearch (C). No 24500353.

Manisera M. van der Kooij A J. Dusseldorp E. 2010. *Identifying the Component Structure of Satisfaction Scales by Nonlinier Principal Component Analysis*. Journal Quality Technology and Quantitative Management. 7(2). pp.97-115.

Michaildis G. De Leuw J. 1998. *The Gifi System of Descriptive Multivariate Analysis. Statistical Science.13(4):307-336*.

Michaildis G. De Leuw J. *The Gifi System of Descriptive Multivariate Analysis. Los Angeles*; International Program in Statistics. University of California.

Safitri D. Widiharih T. Wilandari Y. Saputra A H. 2012. *Analisis Cluster pada Kabupaten/Kota di Jawa Tengah* *berdasarkan Produksi Palawija*. 5(1):11-16.

Van De Burg E. De Leuw J. 1988. *Homogenity Analysis with k Sets of Variables; An Alternating Least Square Method with Optimal Scaling Features*. Netherlands: University of Twente and Leiden University. The Psycometric Society. 53(2). 177-197.

Waryanto B. Millafati Y A.2006. *Transformasi Data Skala Ordinal ke Interval menggunakan Makro Minitab*. Informatika Pertanian. 15:881-895.

Yunianto Y. 2011. *Perbandingan Metode Penggerombolan dengan Komponen Utama Nonlinier dan Gerombol Dua Langkah pada Data Campuran* [skripsi]. Bogor(ID): Institut Pertanian Bogor.