PROCEEDING ISBN : 978 – 979 – 16353 – 7 – 0

P – 14

Factor Analysis Of Ordinal Data Based On Weighted Ranking And Its Application To Reduce Perception Variables To Math Lessons Of Senior High School Student

Kariyam & Perdana, R.B.

Department of Statistics
Islamic University of Indonesia
e-mail: kariyam@uii.ac.id

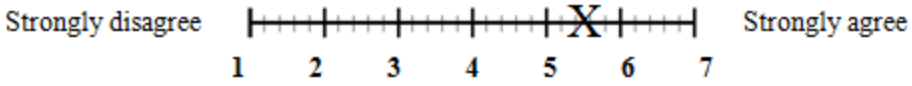
Abstract

The objectives of research are to reduce dimension of ordinal data using factor analysis based on weighted ranking correlation. In general, the correlation is used Spearman ranking correlation. This papers will discuss about application of both methods on the case of simplify of student's perception variables on math lessons. The samples of this research are 791 students which consist of questionnaire. Factor analysis based on the weighted ranking correlation have given the results the number of factors and variables domination on factors more good than Spearman ranking correlation. The factors that influence perception of senior high school students towards Math lesson are an internal motivation of students, negative assessment of the Math lesson, teaching methods of Math teacher, habit of learning of students in Math and the support of parents, and knowledge of impact and benefits of Math.

Key Words: factor analysis, weighted ranking, perception, math

I. INTRODUCTION

Collection of data using questionnaire that contains some statements with closed answers in a form multiple choices, often ease for respondent to fill the questionnaire. Prayudi and Kariyam, (2009) have examined the measurement models E-learning Readness Index (eLRI) as a tool of evaluating the implementation of e - learning in the UII. This indexes is calculated based on the data set that includes the type of metrics, and to collect questionnaire used data. Answer of the questionnaire was designed such that the data gathered, will have an interval scale, and respondents can provide an assessment on a range of value. One of example statement in the indicator, that is "How does the assessment about the seriousness of the students in conducting with e-learning system" then the answers given by the following forms:



One difficulty of answers on the questionnaire above, if done manually, is read correctly from the answers of respondents who replied by giving a cross at any point

within the interval. Suppose that the respondents provide an assessment as showed in the figure above, and then the answer in question is 5.6. Organization of data involving respondents in large numbers by using the model answers on the questionnaire, likely to lead to the occurrence of errors, especially when entering data into the computer. Questionnaires with these forms is to be made easier when a computer-based systems, where the respondent is using the mouse to draw a line to a point corresponding assessment. The touch of the electronic medium is becoming increasingly expensive and difficult, especially if respondents are not accustomed to using computers. This means that the questionnaire model is not flexible to be applied.

Base on the advantages and disadvantages of the model questionnaire above, preferably uncommon use the model questionnaire with multiple choice answers are ordinal type. For example the study aimed to determine the group factor in the process of high school students perceived learning mathematics judged satisfactory, then the questionnaire can be presented in the closed answer as follows:

Table 1. Example for Questionnaire

No.	Statements	Assessment			
		Strongly Agree	Agree	Disagree	Strongly disagree
1.	Math is a subject that I like				
2.	Learning Math is fun				
3.	I never study mathematics, because mathematics is a difficult lesson				
4.	If there are additional hours of math, I do not want to follow				
5.	Teacher of Math uses the computer for teaching so I was motivated to learn math				

Presentation in another form by asking respondents to give a single answer choice, with a similar meaning, for example “The way of teaching can to prod me to study Math better”, with model answers as follows:

a. Strongly agree
b. Agree

c. Disagree

d. Strongly disagree

The advantages of both forms of writing of the questionnaire are easy to answer, because the choices are a direct answer, so it is relatively easier for the respondent. Moreover, the process of gathering and data entry into a computer had a smaller chance of error. Thus both the respondent and researcher are relatively simplified by a questionnaire. Some facts gave the reason that these questionnaires presenting models above more preferable than to assign a value in any point on the range of interval. Disadvantage when possessed as categorical data set is limited in choosing the method of analysis.

Base on the background above in this paper will be discussed factor analysis for data sets with ordinal scale based on weighted ranking the data. Furthermore, this method will be applied to reduce the dimensions of the variables relating to the assessment of high school students on the subjects of Mathematics. The purpose of this study is to provide another alternative factor analysis in particular with respect to ordinal data. Benefit from the application of this method for reduction variable dimension with respect to the perception students towards mathematics courses is to know the groups of variables which assessed students is satisfactory. The results are expected to provide input for the school, teacher with fields of study of math, and parents to continuous improvement.

II. RESEARCH METHOD

This article is part of a research by author about the distance measure based on the weighted of pattern structure and entropy in a non-metric data type object. In this article will be discussing about the use of ordinal data transformation methods that have been proposed by Kauffman and Rousseauw as an intermediate step in the process of factor analysis. This result will be compared with factor analysis using Spearman ranking correlation. Furthermore, both methods will be applied if of high school students' perceptions of math lesson. The factor analysis procedure for ordinal data is to

determine the weighted of the pattern structure, and then used to calculate the correlation matrix as the basis of factor analysis. An outline of the procedure in terms of case studies is distributing questionnaires that have been valid and reliable to 791 high school students who are spread across 12 public and private high schools in Magetan district.

Tool of data collection used in this study is a questionnaire enclosed with the model answers as showed in table 1. Indicators of perceptions of high school students toward mathematics lesson outlined in 20 statements, that is Mathematics is a subject that I like (V1), Math is fun (V2), I never want to learn Math because Math is a difficult subject (V3) , If there are additional hours of math, I do not want to follow (V4), I am most happy to work on the problems of Mathematics (V5), My Math notes is never complete (V6), If there is Math homework, I always plagiarize a friend (V7), Math is not used in daily life (V8), Learning Mathematics is very useful for the future (V9), If mastering math, it will be easy to master other subjects (V10), I will learn mathematics just as there will be a test (V11), Math teacher explains clearly (V12), The way of Math teach, can to prod me for study Mathematics better (V13), My Math teacher can provide the motivation to fun with Math (V14), teachers use computer tools at the time Mathematics lessons that can keep me motivated to learn Math (V15), My parents always motivated me to learn Mathematics (V16), My parents always taught me Math when I was studying at home (V17), I often learn Math together with my friends (V18), I would much prefer to learn and do the job instead of playing (V19), and even though I studied hard, I can't understand Math (V20).

Data analysis that used in this study preferred to use factor analysis of ordinal data based on a weighted ranking. The essential purpose of factor analysis is to describe, if possible, the covariance relationship among many variables in terms of a few underlying, but unobservable, random quantities called, factors. Suppose variables can be grouped by their correlations, where all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group.

The observable random vectors X, with p components, has mean μ and covariance, $\sum$. The factor models postulates that X is linearly dependent on a few

unobservable random variable F_1, F_2, F_3, … , F_q called common factors, and p additional sources of variation $\varepsilon_1, \varepsilon_2, \varepsilon_3, ..., \varepsilon_q$ called error or specific factor. Suppose L denote as the matrix of factor loadings, then model in matrix notation is below:

$$(X - \mu) = L F + \varepsilon \qquad (1)$$

The factor loadings can be derived from correlation matrix R, for example Spearman ranking correlation. The principal component of factor analysis of the sample correlation matrix R is specified in terms of its eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \ldots, (\hat{\lambda}_p, \hat{e}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$. Let $q < p$ be the number of common factors. The matrix of estimated factor loadings $\left\{ \tilde{l}_{ij} \right\}$ is given by

$$\tilde{L} = \left[\sqrt{\hat{\lambda}_1} \, . \, \hat{e}_1 \;\middle|\; \sqrt{\hat{\lambda}_2} \, . \, \hat{e}_2 \;\middle|\; \ldots\ldots \;\middle|\; \sqrt{\hat{\lambda}_q} \, . \, \hat{e}_q \right] \qquad (2)$$

In special case if q factors that included within analysis much enough, say $q > 2$, then sometimes there are difficulty for interpretation of factors, because variables X_i that explained for q factors. To overcome this case, then done one of rotation there is varimax rotation.

Christopher, Z., 2003, reviewed the agglomeration grouping of the data as rankings, based on the proposal of Kaufman, and Rousseuuw, 1990, which suggests the form of transformation the ordinal type of data through two stages. Suppose that a variable of type of ordinal X_k with M_k categories sequential, then the phase transformation is conducted as follows:

(i) Replace X_{rk} with this rank $v_{rk} \in \{1,2,.....,M_k\}$.

(ii) Transform v_{rk} to unit interval by calculating:

$$Z_{rk} = \frac{v_{rk} - 1}{M_k - 1} \qquad (3)$$

Furthermore, the results of this transformation will be used as a basis for calculating the correlation matrix in factor analysis. In principally this way is uncommon, but it is expected can be used as alternative of validation in factor analysis. This results will be compared with factor analysis using the Spearman correlation matrix.

III. RESULT AND DISCUSSION

The first step in the process of data analysis is to transform the ordinal as unit interval with use the equation (3), and further processed by factor analysis. Conformity all assumptions in factor analysis have been checked. Furthermore, based on minimal eigen value 1, then obtained five factors from 20 variables, as shown in figure 1.

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.397	26.984	26.984	5.397	26.984	26.984	2.761	13.805	13.805
2	1.669	8.347	35.331	1.669	8.347	35.331	2.494	12.471	26.276
3	1.613	8.065	43.396	1.613	8.065	43.396	2.145	10.723	36.999
4	1.336	6.682	50.078	1.336	6.682	50.078	2.057	10.284	47.283
5	1.119	5.597	55.676	1.119	5.597	55.676	1.679	8.393	55.676
6	.964	4.820	60.496						
7	.854	4.269	64.765						
8	.812	4.061	68.826						
9	.763	3.814	72.640						
10	.674	3.371	76.011						
11	.619	3.095	79.106						
12	.610	3.048	82.154						
13	.565	2.823	84.977						
14	.548	2.739	87.716						
15	.508	2.539	90.255						
16	.495	2.475	92.731						
17	.481	2.404	95.135						
18	.428	2.138	97.272						
19	.332	1.662	98.935						
20	.213	1.065	100.000						

Figure 1. Eigen Value and Total Variance Explained

The proportion of the total population variance from rotation that can be explained by the first factor of 13.8%, the second factor 12.5%, 10.7% third factor, the fourth factor 10.2%, and the fifth factor 8.4%. All these factors can explain the total variance of the population 55.7%, as shown in Figure 1. Furthermore, based on the rotated component matrix, then some variables that dominate each factor is as shown in Figure 2. below.

	Component				
	1	2	3	4	5
V1	.867	.050	.163	.072	.103
V2	.807	.060	.216	.114	.137
V5	.710	.182	.144	.272	.053
V7	.311	.637	.056	.150	-.039
V6	-.179	.635	.086	.169	.027
V11	.182	.626	.188	.119	.153
V4	.120	.579	.088	.118	.275
V20	.474	.478	.047	-.181	-.143
V3	.404	.411	-.025	.058	.235
V13	.154	.110	.809	.119	.116
V14	.338	.142	.753	.140	.102
V12	.063	.094	.745	.107	.164
V17	.210	-.005	.030	.757	.038
V16	.111	.087	.065	.724	.156
V18	.021	.289	.102	.534	.158
V15	-.077	.109	.317	.494	-.167
V19	.235	.371	.228	.372	.073
V9	.151	.195	.171	.067	.753
V10	.084	-.034	.142	.171	.685
V8	-.016	.448	.027	-.077	.547

Figure 2. Rotated Component Matrix

Based on the rotated component matrix, then variables that dominated each factor are as follows:

a. The first factor dominated by three variables, that is Mathematics is a subject that I like (V1), Math is fun (V2), and I am most happy to work on the problems of Mathematics (V5). Further, this factor called internal motivation of students.
b. The second factor dominated by six variables, that is I never want to learn Math because Math is a difficult subject (V3), If there are additional hours of math, I do not want to follow (V4), My Math notes is never complete (V6), If there is Math homework, I always plagiarize a friend (V7), I will learn mathematics just as there will be a test (V11), and even though I studied hard, I can't understand Math (V20). This factor is called as negative assessment of the Math lessons.
c. The third factor dominated by three variables, that is Math teacher explains clearly (V12), the way of Math teach, can to prod learning of Math better (V13), My Math teacher can provide the motivation to fun with Math (V14). This factor called as teaching methods of Math teacher.
d. The fourth factor dominated by five variables, that is teachers use computer tools at Math lessons that can keep me motivated to learn Math (V15), My parents always motivated me to learn Math (V16), My parents always taught me Math

when I was studying at home (V17), I often learn Math together with my friends (V18), I would much prefer to learn and do the job instead of playing (V19). This factor called habit of learning of student in Math and the support of parents.

e. The fifth factor dominated by three variables, that is Math is not used in daily life (V8), Learning of Math is very useful for the future (V9), and, if mastering Math, it will be easy to master other subjects (V10). This factor called as knowledge of impact and benefits of Math.

The result of data analysis using the factor analysis based on the Spearman ranking correlation as below:

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Var 1	-0.633	0.608	0.014	-0.028	0.159
Var 2	-0.655	0.528	0.054	0.001	0.151
Var 3	-0.514	0.092	-0.321	-0.065	0.105
Var 4	-0.529	-0.232	-0.333	-0.013	-0.033
Var 5	-0.667	0.377	0.045	-0.190	0.102
Var 6	-0.335	-0.433	-0.277	-0.148	-0.272
Var 7	-0.563	-0.013	-0.320	-0.257	-0.211
Var 8	-0.372	-0.302	-0.428	0.261	0.161
Var 9	-0.520	-0.212	-0.165	0.396	0.403
Var 10	-0.373	-0.207	0.043	0.340	0.476
Var 11	-0.593	-0.165	-0.294	-0.043	-0.174
Var 12	-0.504	-0.108	0.330	0.403	-0.260
Var 13	-0.580	-0.034	0.363	0.390	-0.311
Var 14	-0.676	0.101	0.316	0.307	-0.263
Var 15	-0.291	-0.254	0.387	-0.220	-0.206
Var 16	-0.463	-0.275	0.322	-0.343	0.250
Var 17	-0.430	-0.140	0.400	-0.439	0.251
Var 18	-0.459	-0.342	0.113	-0.239	0.084
Var 19	-0.583	-0.115	0.040	-0.167	-0.071
Var 20	-0.407	0.332	-0.385	-0.111	-0.268
Variance	5.3968	1.6694	1.6131	1.3364	1.1195
% Var	0.270	0.083	0.081	0.067	0.056

Cumulative percent of variance that explained by the fifth factors is 55.7%, and its value the same with factor analysis based on the weighted ranking correlation. Even though, relative difficult to interpretation of each factor because majority variables have been dominated in the first factor. The number of factors if doing by rotation is 13 factors, and of course this value is bigger. So, this result is not better than factor analysis based on the weighted ranking correlation.

More detail of percentage observations to the achievements of high school students who expressed at least agree to the statement about the perceptions of students towards Math lesson, as shown in Figure 3. below.

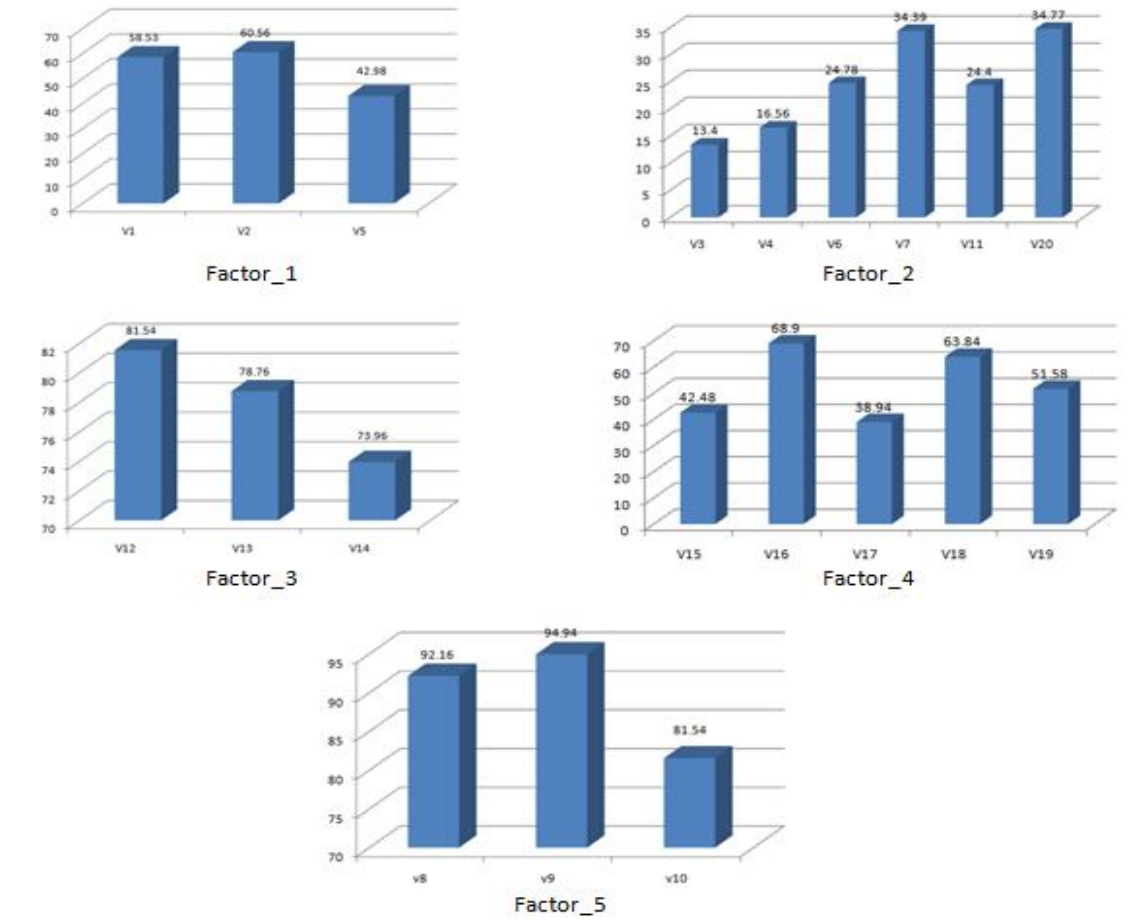


Figure 3. Percentage of High School Students with a Minimum Agree Perception

The student with the perception that Math is a subject that they like and Math is fun around 60%. Whereas, the student with statement that they most happy to work on the problem of Math are around 43%. It is showed that the student have relatively a good internal motivation. Majority of the student disagree with contents of variables in the second factor. Around 83% of students disagree with statements that they never want to learn Math because Math is a difficult and they do not follow to additional hours of Math. Similarly, around 65% of students disagree with statement that they can't understand Math even though they studied hard. Meanwhile, around 75% of students have a complete of notes of Math and they will learn Math not only when they have exam. Around 66% of students not plagiarize a friend when they have Math homework. In general, majority of students disagree with the negative assessment of the Math lessons. Nevertheless, the habit of plagiarize a friend for Math homework to be apprehensive, except in context discussion. At least 74% of students have a good perception about teaching methods of Math teacher. The students agree that the Math teacher have explained clearly and Math teacher can provide the motivation to learn of Math better. Majority of Math teacher not use computer as tools to teach. Around 39%

of parents have taught Math at home, and 69% of parents have given support to study Math. Around 64% of students learning habit are together with their friends. Majority of students have a good perception about knowledge effect and benefit of Math. Percentage showed it of students that agree and strongly agree to statements learning of Math is very useful for the future and if mastering Math, it will be easy to master the other subjects, accomplished more than 80%. In general, the students of high school in Magetan have a good perception about Math lessons.

IV. CONCLUSION AND SUGGESTION

The weighted ranking pattern can be used to intermediate step in factor analysis. Application this method to case of perception of senior high school student showed a good result than factor analysis based on the Spearman ranking correlation. The factors that influence perception of senior high school students towards Math lessons are an internal motivation, assessment to Math lesson, teaching method of Math teacher, habit of learning of students on Math lesson, and support of parents, and knowledge of impact and benefits of Math.

V. REFERENCES

Christoper, Z., 2003, *Agglomerative Clustering of Rankings Data, with an Application to Prison Rodeo Events*, http://polmeth.wustl.edu/media/Paper/zorn03.pdf

Deng, S., Xu, X., dan He, Z., (2006), *Improving Categorical Data Clustering Algorithm by Weighting Uncommon Attribute Value Matches*, Journal of ComSIS Vol.3, No.1, June 2006.

Finch, H., 2005, *Comparison of Distance Measures in Cluster Analysis with Dichotomous Data*, Journal of Data Science, Vol. 3, 85 – 100, 2005.

Johnson, R.A., and Wichern, D.W., (1992), *Applied Multivariate Statistical Analysis (3nd ed)*, New Jersey : Prentice Hall

Kariyam, et al, 2009, *Analysis of student satisfaction at the senior high school*, Islamic University of Indonesia, unpublished.

Kaufman, L., and Rousseeuw, P.J., 1990, *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: John Wiley & Sons, Inc.

Rezankova, H., 2010, *Cluster Analysis and Categorical Data*, http://panda.hyperlink.cz/cestapdf/pdf09c3/rezankova.pdf.

Perdana, R.B., and Kariyam, 2011, *Perception of Student of Senior High School to Math Lessons*, research report, unpublished.

Prayudi, and Kariyam, 2010, The measurement models of E-learning Readness Index (eLRI), Islamic University of Indonesia, research report, unpublished.