Preference Mapping Selection Of Smartphone's Brands And Prices Based On Allowance Using Multiple Correspondence Analysis

Maulida Rahmatul Husna*, Suci Astutik**
* Statistics Program, Departement of Mathematics, Brawijaya University Malang
** Statistics Program, Departement of Mathematics, Brawijaya University

Article Info

Article history:

Received Jul 12th, 2017
Revised Aug 20th, 2017
Accepted Oct 26th, 2017

Keyword:

Correspondence Analysis
Multiple correspondence
Smartphones
Mapping
Allowance

ABSTRACT

The use of smartphone plays an important role for the community and among students, particularly students of the University of Brawijaya Malang where the needs of information increasingly dominate. The needs of the student can be affected by student's allowance, because the majority of students have additional allowance. The mapping can be done to obtain information about the preferences selection brand smartphone and smartphone based on the price of the allowance. Statistical methods that can be used for mapping is Multiple Correspondence Analysis (MCA). The purpose of this research is to apply for MCA selection preference mapping brand smartphone and smartphone's price based of the allowance so that the information obtained by the relationship of allowance against the selection preferences of smartphones. The research using primary data with purposive technique sampling that comes from the spread of the questionnaire with respondents was University students of Brawijaya Malang batch 2015. With MCA acquired a two-dimensional graph with the diversity proportion of 76,103%, there are three cluster shaped: obtained the information of students who got high allowance have preference to choose two type of price Rp 2.500.000 – Rp 3.000.000 or mote than Rp 3.000.000 the preference selected brand are Iphone, Samsung or Sony. Medium Allowance range have a preference of price Rp 2.000.000 – Rp 2.500.000, preference of selected brand were Vivo, Lenovo, Oppo or Asus, and low allowance have two type price preference Rp 1.500.000 or Rp 1.500.000 – Rp 2.000.000. LG, Smartfren Microsoft, Xiaomi or Advan are selected preference brand.

Corresponding Author:

Suci Astutik,
Statistics Program,
Departement of Mathematics,
Brawijaya University Malang,
Jl. Veteran No. 05 Malang, Jawa Timur, Indonesia 65144
Email: suci_sp@ub.ac.id

1. INTRODUCTION

At the time of the increasingly developing science, technology and innovation to support the present needs of the community. Innovations in technology one exists on the means of communication are now presented in the form of a more modern and practical form of mobile or cell phone and innovate into smartphones. The use of smartphones has an important role among the students, especially students of the University of Brawijaya Malang where the needs of information increasingly dominate.

Smartphone purchasing decision is affected by lifestyle and reference groups. Purchasing behavior also influenced by personal factors, which refers to work related to income, economic environment, and the nature or personality. On among the students, referring to the factor income because most students still rely on the income of the parents, then allowance can be linked to smartphone selection preferences.

Mapping can display the position of the coordinate axis to detect and provide an explanation of the relationship between two or more categorical variables. Statistical methods that can be used for mapping is a Multiple Correspondence Analysis (MCA). MCA has input in the form of a matrix of indicators from a contingency table is demonstrated through the rows and columns and involving two or more categorical variables,

as well as give an output in the form of graphs. Therefore, based on the description above can be carried out research entitled "Preference Mapping Selection of Brands and Prices of Smartphones Based Allowance Using Multiple Correspondence Analysis"

2. RESEARCH METHOD

a. Correspondence Analysis

Correspondence analysis is a multivariate data exploration technique on the relationship between two or more variables by simultaneously modeling rows and columns based on a two-way contingency table in a low-dimensional vector space [1]. There are two kinds of correspondence analysis according to type of variables, simple correspondence analysis and multiple correspondence analysis. The method for analyzing contingency tables with two variables called simple correspondence analysis and the method for analyzing contingency tables of more than two variables is called multiple correspondence analysis. Correspondence analysis has several basic properties:

a. Applied to non-metric data on nominal and ordinal scale measurement.
b. Describes non-linear relationships.
c. There is no assumption about distribution.
d. Models are not hypothesized.
e. A multivariate method whose final result can be a hypothesis that may need to be tested further.
f. One method of grouping variables in addition to cluster analysis is also used for dimensional reduction.

The purpose of correspondence analysis is:

a. Compare the similarity of two categories form first variable (line) based on the second qualitative variable (column).
b. Know relationship between one category of row variable with one column category.
c. Describes each variables category of row and column according to contingency table into the correspondence graph, so that it can displayed simultaneously in a small dimension vector space optimally.

The concept in correspondence analysis is singular value decomposition. Correspondence analysis illustrates the proximity of profiles between categories on each indicator in the form of a two-dimensional graph. Interpretation of the correspondence analysis output results by looking at the value of inertia that shows the major axis contribution to total inertia [2]. If two main axes provide a sufficiently large inertia value, then it can be interpreted that the main axis is quite representative the information.

Calculating the inertial axis is done by squaring the singular value. The singular value is sum of squared plots to the center of axis which weighted by the mass of each plot. The plot contribution to root character or to the major axis is called the absolute contribution. Absolute contributions show proportion of diversity that each category describes on each axis. Thus, plots with mass values greater or further distant from the center of the axis can contribute more and more informatively.

b. Multiple Correspondence Analysis

Multiple Correspondence Analysis (MCA) is performed on a multi-directional contingency table or involves more than two categorical variables and one variable may still have multiple levels. If MCA involves 3 variables, variable 1 as a row consisting of i category and variable 2 as rows consisting of j category and varible 3 as columns consisting of k category, shown in Table 1.

Table 1. Coningency Table with 3 Variables

Peubah 1	Peubah 2	Peubah 3					Total
		1	2	3	⋯	c	
1	1	n_{111}	n_{112}	n_{113}	⋯	n_{11c}	$n_{11.}$
	2	n_{121}	n_{122}	n_{123}	⋯	n_{12c}	$n_{12.}$
	3	n_{131}	n_{132}	n_{133}	⋯	n_{13c}	$n_{13.}$
	⋮	⋮	⋮	⋮	⋱	⋮	
	r'	$n_{1r'1}$	$n_{1r'2}$	$n_{1r'3}$	⋯	$n_{1r'c}$	$n_{1r'.}$
Total		$n_{1.1}$	$n_{1.2}$	$n_{1.3}$		$n_{1.c}$	$n_{1..}$
2	1	n_{211}	n_{212}	n_{213}	⋯	n_{21c}	$n_{21.}$
	2	n_{221}	n_{222}	n_{223}	⋯	n_{22c}	$n_{22.}$

	3	n_{231}	n_{232}	n_{233}	⋯	n_{23c}	$n_{23.}$
	⋮	⋮	⋮	⋮	⋯	⋮	
	r'	$n_{2r'1}$	$n_{2r'2}$	$n_{2r'3}$	⋯	$n_{2r'c}$	$n_{2r'.}$
Total		$n_{2.1}$	$n_{2.2}$	$n_{2.3}$		$n_{2.c}$	$n_{2..}$
3	1	n_{311}	n_{312}	n_{313}	⋯	n_{31c}	$n_{31.}$
	2	n_{321}	n_{322}	n_{323}	⋯	n_{32c}	$n_{32.}$
	3	n_{331}	n_{332}	n_{333}	⋯	n_{33c}	$n_{33.}$
	⋮	⋮	⋮	⋮	⋯	⋮	⋮
	r'	$n_{3r'1}$	$n_{3r'2}$	$n_{3r'3}$	⋯	$n_{3r'c}$	$n_{3r'.}$
Total		$n_{3.1}$	$n_{3.2}$	$n_{3.3}$		$n_{3.c}$	$n_{3..}$
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮
	⋮	⋮	⋮	⋮	⋯	⋮	⋮
	⋮	⋮	⋮	⋮	⋯	⋮	⋮
	⋮	⋮	⋮	⋮	⋱	⋮	⋮
	⋮	⋮	⋮	⋮	⋯	⋮	⋮
r	1	n_{r11}	n_{r12}	n_{r13}	⋯	n_{r1c}	$n_{r1.}$
	2	n_{r21}	n_{r22}	n_{r23}	⋯	n_{r2c}	$n_{r2.}$
	3	n_{r31}	n_{r32}	n_{r33}	⋯	n_{r3c}	$n_{r3.}$
	⋮	⋮	⋮	⋮	⋯	⋮	⋮
r	r'	$n_{rr'1}$	$n_{rr'2}$	$n_{rr'3}$	⋯	$n_{rr'c}$	$n_{rr'.}$
Total		$n_{r.1}$	$n_{r.2}$	$n_{r.3}$	⋯	$n_{r.c}$	$n_{r..}$

Where, $i = 1,2, \dots, r$; $j = 1,2, \dots, r'$ and $k = 1,2, \dots, c$

To know the relation of closeness between two category variables in contingency table is use Pearson Chi-Squared test with hypothesis:

H_0: $\rho = 0$ vs

H_1: $\rho \neq 0$

Using the test statistics:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-m_{ij})^2}{m_{ij}} \quad (1)$$

$$m_{ij} = \frac{n_{i.}\times n_{j}}{n_{..}} \quad (2)$$

MCA analyzes the data set containing dummy variable called the indicator matrix. Indicator matrix is the transformed matrix of contingency table containing the dummy variable. Suppose the indicator matrix is denoted as $\boldsymbol{N}$ with r row and c column. If there is w category variables, $\boldsymbol{N}$ is the indicator matrix of the category variables of w. Thus, is the combined indicator matrix of all variables. The indicator matrix is described in equation below:

$$\boldsymbol{N_I} = \begin{bmatrix} n_{I11} & n_{I12} & \dots & n_{I1c} \\ n_{I21} & n_{I22} & \dots & n_{I2c} \\ \vdots & \vdots & \ddots & \vdots \\ n_{Ir1} & n_{Ir2} & \dots & n_{Irc} \end{bmatrix} \quad (3)$$

$$\boldsymbol{N_{II}} = \begin{bmatrix} n_{II11} & n_{II12} & \dots & n_{II1c} \\ n_{II21} & n_{II22} & \dots & n_{II2c} \\ \vdots & \vdots & \ddots & \vdots \\ n_{IIr1} & n_{IIr2} & \dots & n_{IIrc} \end{bmatrix} \quad (4)$$

$\vdots \qquad\qquad \vdots$

$$\boldsymbol{N}_w = \begin{bmatrix} n_{w11} & n_{w12} & \dots & n_{w1c} \\ n_{w21} & n_{w22} & \dots & n_{w2c} \\ \vdots & \vdots & \ddots & \vdots \\ n_{wr1} & n_{wr2} & \dots & n_{wrc} \end{bmatrix} \quad (5)$$

$$\boldsymbol{N} = [\boldsymbol{N}_I \quad \boldsymbol{N}_{II} \quad \dots \quad \boldsymbol{N}_w] \quad (6)$$

Indicator matrix then transformed into a Burt matrix, by transposing the indicator matrix and multiplied by itself. Suppose the indicator matrix denoted as $\boldsymbol{N}$, then the Burt matrix denoted as $\boldsymbol{B}$ and defined as $\boldsymbol{N'N}$.

$$\boldsymbol{B} = \boldsymbol{N'N} = \begin{bmatrix} \boldsymbol{N}_I' \\ \boldsymbol{N}_{II}' \\ \vdots \\ \boldsymbol{N}_w' \end{bmatrix} [\boldsymbol{N}_I \quad \boldsymbol{N}_{II} \quad \dots \quad \boldsymbol{N}_w] = \begin{bmatrix} \boldsymbol{N}_I'\boldsymbol{N}_I & \boldsymbol{N}_I'\boldsymbol{N}_{II} & \dots & \boldsymbol{N}_I'\boldsymbol{N}_w \\ \boldsymbol{N}_{II}'\boldsymbol{N}_I & \boldsymbol{N}_{II}'\boldsymbol{N}_{II} & \dots & \boldsymbol{N}_{II}'\boldsymbol{N}_w \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{N}_w'\boldsymbol{N}_I & \boldsymbol{N}_w'\boldsymbol{N}_{II} & \dots & \boldsymbol{N}_w'\boldsymbol{N}_w \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1c} \\ b_{21} & b_{22} & \dots & b_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \dots & b_{rc} \end{bmatrix} \quad (7)$$

Suppose a correspondence matrix with $r \times c$ data, a visualization of the contingency table in the form of matrix's rows and columns in low dimension by constructing matrix $\boldsymbol{P}_{(r\times c)}$ as the correspondence matrix [3].

$$\boldsymbol{P} = \frac{\mathbf{1}}{b_{..}}\boldsymbol{B} = \frac{1}{b_{..}} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1b} \\ b_{21} & b_{22} & \dots & b_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ b_{a1} & b_{a2} & \dots & b_{ab} \end{bmatrix} \quad (8)$$

to reduce the dimensions of data based on the largest data diversity and maintain the optimum information required Singular Value Decompotition (SVD). SVD is concept of matrix algebra and eigen decomposition concept. SVD results are eigenvalues (λ) and eigenvectors u_i and v_i. $\boldsymbol{Z} = \boldsymbol{U\Lambda V'}$

Provided that $\boldsymbol{U'U} = \boldsymbol{V'V} = \boldsymbol{I}$, $\boldsymbol{U}$ is an orthogonal matrix in which the elements are eigenvectors of $\boldsymbol{PP'}$ and $\boldsymbol{V}$ is an orthogonal matrix in which the elements are eigenvectors of $\boldsymbol{P'P}$. The matrix $\boldsymbol{\Lambda}$ is the diagonal matrix $n \times n$ whose diagonal is the eigenvalue $(\lambda_1, \lambda_2, \dots \lambda_b)$. Coordinate columns ($\boldsymbol{X}$) and row coordinates ($\boldsymbol{Y}$) are defined as follows:

$$\boldsymbol{X} = \boldsymbol{D_c^{-1}D_c^{1/2}V\Lambda} \qquad \boldsymbol{Y} = \boldsymbol{D_r^{-1}D_r^{1/2}U\Lambda} \quad (12)$$

c. Correspondence Graph

Output of MCA is a graph named correspondence graph, the correspondence graph is presented on a symmetry map and Chi-Square distance representing the similarity between categories. Illustrated by graph in which row and column coordinates are standardized to obtain orthogonal dimensions that can place each category in best position [4].

The plotting of correspondence graph illustrates by two best dimensions for representing data at point coordinates that measure the value of information for each dimension, by looking at the value of inertia to indicate the major axis contribution to total inertia. If two main axes provide a sufficiently large inertia value, then it can be interpreted that the main axis is quite representative of information. The distance between coordinate points is meaningless, but the coordinate points in the same quadrant or around adjacent coordinates indicate relationship. If two adjacent coordinates represent each categories of the same variable then two categories can be categorized [3].

To illustrate the coordinates, each category of row and column variables are given different symbols or traits (eg. different colors or different letter colors). Correspondence graph is illustrated in Figure 1.

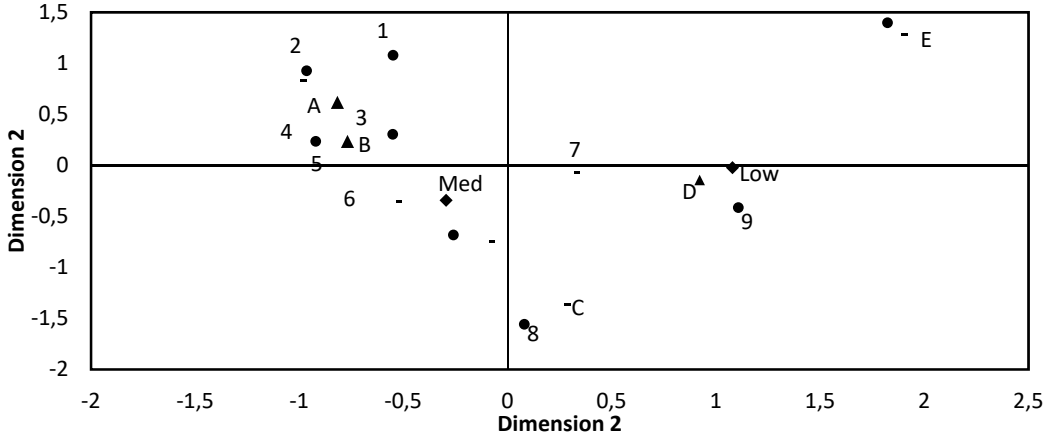

Figure 1. Ilustration of Correspondence Graph

d. Research Data

Primary data used in this study, obtained through questionnaire. Respondents of this study are active bachelor study of Brawijaya University, force 2015. Primary data collection is influenced by population, sample and sampling technique. The sampling technique is expected to give the results of research with a high level of accuracy and minimal expenditure [5]. Population in this research is smartphone user and must be active bachelor

study of Universitas Brawijaya Malang force 2015, the number of population can not known for certain (infinite population).
Sampling technique by population type is divided into two, namely probability sampling and nonprobability sampling. This research used nonprobability sampling as technique, that is sampling technique based on consideration determined by researcher [6]. Nonprobability sampling is used for infinite population.
There are several sampling methods on nonprobability sampling technique, such as systematic sampling, quota sampling, accidental sampling, purposive sampling, saturated sampling and snowball sampling. The method used in this research is purposive sampling that based on certain criterion which have been determined by researcher. Sample criteria used in this study are:
1. Active students of Universitas Brawijaya Malang force 2015 and smartphone users.
2. Active students of Universitas Brawijaya Malang force 2015 who choose their own smartphone used.
3. Active students of Universitas Brawijaya Malang force 2015 smartphone users who each month get an allowance from parents.

The formula for calculating the number of samples with the study population corresponds to the distribution of Bernoulli where p and q are the proportion of male or female respondents selected [7], as follows:

$$n = \frac{(Z_{\alpha/2})^2 pq}{\epsilon^2} = \frac{1.96^2 \times 0.5 \times 05}{0.1^2} = 96.04 \approx 97$$

e. Steps of Analysis

The steps of data analysis in this study are (1) compile the research questionnaire (2) collect the data (3) create the contingency table (4) test the freedom between categories (5) form the matrix N (6) form the matrix B (7) form the correspondence matrix (8) calculates the singular value (9) calculates the coordinate axis of the main row and column (10) forming the correspondence graph (11) interpretation graph

3. RESULTS AND ANALYSIS

3.1. Freedom Test

Prior to conducting correspondence analysis, a Pearson Khi-squared test was performed to determine the relationship between the variables. The contingency table can be seen in Appendix 5. The results of the Pear son Khi-squared test in Table 2.

Table 2. Result ofPearson Chi-square Test for Each Variables

Peubah	Nilai Khi-kuadrat Pearson	P-Value
Merek *smartphone* dengan harga *smartphone*	110.707	0.000
Merek *smartphone* dengan uang saku	36.875	0.045
Harga *smartphone* dengan uang saku	205.468	0.000

Hypotheses:
H_0: $\rho_1 = \rho_2 = 0$
H_1: $\rho_1 = \rho_2 \neq 0$

From Table 1 it is found that the p-value for Pearson's Chi-squared test between smartphone brand variables with smartphone's price, smartphone brand with allowance, and smartphone price with allowance less than 0.05 so that it decided to reject H_0. This means there are relationship between smartphone's brands with smartphone's prices, smartphone's brands with allowance and smartphone's prices with allowance. Because there are relationship between variables, the MCA can be applied.

3.2. Multiple Correspondence Analysis (MCA)

Correspondence analysis is performed to determine any relationship between two variables. MCA involving more than 2 category variables. MCA begins by calculating the eigenvalues or the value of the primary inertia using SVD, a coefficient expressing the proportion of diversity described by the dimension. The eigenvalues along with the cumulative percentages are presented in Table 3.

Table 3. shows that the eigenvalue for dimension 1 is 0.378 with the percentage of eigenvalues is 61.137%. So it means that the proportion of data diversity can explained as 61.137% with 1 dimensional mapping. In second dimension obtained eigen value 0.091 or 14.966%, meaning that in dimension 2 data diversity can be explained as 14.966%. So with two dimensional mapping, proportion of data diversity can be explained is 76.103% according to the cumulative percentage of eigenvalues by involving the second dimension.

Table 3. Eigen Value and Percentage Cumulative Eigen Value

Dimension	Eigen Value	Percentage Eigen Value	Cumulative Percentage Eigen Value
Dimension 1	0.378	61.137	61.137
Dimension 2	0.091	14.966	76.103
Dimension 3	0.035	5.735	81.838
Dimension 4	0.025	4.124	85.926
Dimension 5	0.008	1.271	87.233
Dimension 6	0.000	0.035	87.268
Dimension 7	0.000	0.000	87.268

The coordinates of variable category also named column profile coordinates are main coordinates for each variable's category to determine the position of each category in subspace each dimension. The coordinates of column profile are described in Table 4. In Table 4 we get location information of the category variables in a graph, dimension 1 or the first major axis is used as axis, and dimension 2 or second main axis is used as the ordinate.

Table 4. Coordinate Profil Column

Variables	Category of Variables	Dimension 1	Dimension 2
Brands of Smartphones	Advan	-0.777	0.196
	Asus	0.079	-1.557
	LG	-0.550	1.080
	Lenovo	-0.261	-0.681
	Oppo	-0.090	-0.751
	Samsung	0.319	-0.071
	Sony	1.107	-0.414
	Vivo	-0.534	-0.358
	Xiaomi	-0.922	0.236
	iPhone	1.823	1.397
	Microsoft	-0.552	0.305
	Smartfren	-0.966	0.928
Prices of Smartphones	Very Low (>Rp 1.500.000)	-0.817	0.616
	Low (Rp 1500.000 – Rp 2.000.000)	-0.768	0.233
	Medium (Rp 2.000.000 – Rp 2.500.000)	0.273	-1.362
	High (Rp. 2.500.000 – Rp. 3.000.000)	0.922	-0.145
	Very High (>Rp. 3.000.000)	1.281	1.281
Month Allowance	Low (<Rp1.000.000)	-0.995	0.828
	Medium (Rp 1.000.000 - Rp 2.000.000)	-0.295	-0.344
	High (>Rp 2.000.000)	1.079	-0.024

The distance between coordinate points is meaningless, but the coordinate points in the same quadrant or around adjacent coordinates indicated relation [2]. If two adjacent coordinates represent the categories of the same variable then two categories can be categorized. Corespondence graph is presented in Figure 2.

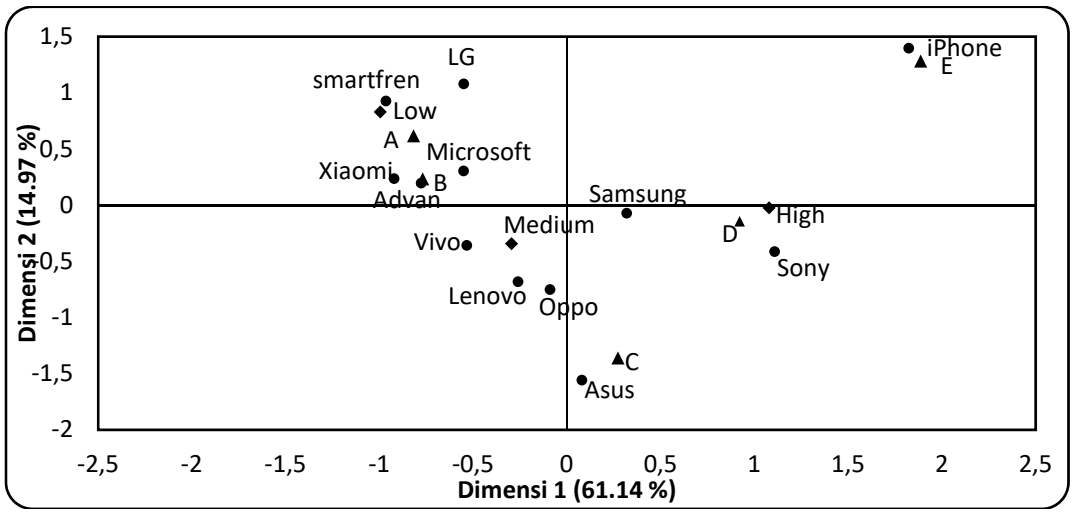

Figure 2. MCA Graph of Brand, Prices and Allowance

a. Dimension 1 (vertical axis) provides information on data diversity of 61.14%, gives an overview of student allowance rate, where the coordinates on the right (positive) indicates higher student allowance rate, while on the left (negative) are lower. Dimension 2 (horizontal axis) provides 14.97% data diversity information. By involving dimension 2 data diversity can be more informative.
b. From Figure 1 Coordinate point of the low allowance category (<Rp 1,000,000) adjacent to price A (smartphone price <Rp 1.500.000) and brands of smartphone's are Smartfren and LG. Thus, students who earn an allowance less than Rp 1,000,000 have a preference Choose Smartfren or LG brand for price less than Rp 1,500,000.
c. The coordinate points for medium allowance categories (Rp 1,000,000 - Rp 2,000,000) are close to Advan, Vivo and Lenovo brands also close to C price (smartphone range price Rp 2,000,000 - Rp 2,500,000). So, students who get an allowance range of Rp 1,000,000 - Rp 2,000,000 have preferences to choose brand Advan, Vivo or Lenovo brands with price range Rp 2,000,000 - Rp 2,500,000.
d. The coordinate point of the high allowance category (> Rp 2,000,000) is close to the D price (smartphone price range Rp 2,500,000 - Rp 3,000,000), Samsung and Sony brands. Thus, students who get allowance of more than Rp 2,000,000 have a preference for choosing a Samsung or Sony brand with a price range of Rp 2,500,000 - Rp 3,000,000.

4. CONCLUSION

The conclusions of this research are:

1. The force 2015 students with high allowance tend have two preference smartphone price category, which are Rp 2,500,000 - Rp 3,000,000 or more than Rp 3,000,000. Preferred choosed brand preference Iphone, Samsung or Sony.
2. The force 2015 students with medium allowance tend have smartphone price preference Rp 2,000,000 - Rp 2,500,000. Preferred brand preference is Vivo, Lenovo, Oppo or Asus.
3. The force 2015 students with low allowance tend to have a smartphone preference with two price categories, which is less than Rp 1,500,000 or Rp 1,500,000 - Rp 2,000,000. LG, Microsoft Smartfren, Xiaomi or Advan are preferred brand preference.

ACKNOWLEDGEMENTS

I would to thank the Ministry of Research, Technology, and Higher Education of Republic Indonesia for the grants funding fundamental/ basic research flagship college that given to this research.

REFERENCES

[1] Greenacre, M. 1984. Theory and Application of Correspondence Analysis Fifth Edittion. Prentice Hall. New Jersey
[2] Johnson, R.A. dan Winchern, D.W. 2007. Applied Multivariate Statistical Analysis Sixth Edition. Practice Hall International. New Jersey.
[3] Rencher, C.A. 2002. Methods of Multivariate Analysis Second Edittion. A John Wiley and Sons, Inc. Publication. Canada.

[4] Hair, J.F., Black, W.C., Babin, B.J. dan Anderson, R.E. 2010. Multivariate Data Analysis Seventh Edition. Prentice Hall International. New York.
[5] Supranto, J. 1992. Teknik Sampling Statistika – Sampling untuk Pemeriksaan. UI Press. Jakarta.
[6] Nugroho, W.H. 1994. Teknik Penarikan Sampel: teori dan Aplikasi. First Edition. Malang: IKIP.
[7] Walpole, E.R. 1992. Pengantar Statistika Third Edition. PT. Gramedia Pustaka Utama. Jakarta.