

# **The Influence of Consumer Behavior on Purchase Decision on Regression-based Model: Application to Portable Kitchen**

**Muhammad Reza Do. Bagus**

Information System Department

School of Information Management and Computer Tidore Mandiri (STMIK Tidore Mandiri)

Jl. Sultan Zainal Abidin Syah No. 25, Kota Tidore Kepulauan, North Mollucas Province, 97812,

Indonesia

[dobagusmuhammadreza@gmail.com](mailto:dobagusmuhammadreza@gmail.com)

## **Abstract**

Consumer awareness of product design is getting higher and higher. Products exterior design is one of the most important factors which affects consumer's purchasing decisions. Therefore, how to develop the product exterior design which satisfies consumers' concerns and interests effectively is great importance to exterior design manufacturers. To understand contributing factors of products exterior for consumer's purchasing decision, this paper proposed to use Regression-based model. Regression-based model is used to determine what features of product exterior design which attracts consumers' concerns. The target of a Regression-based model is to obtain importance scores that represent factors of products exterior design, hence based on its score, it can be ranked which of the attributes of products exterior design more attract consumers' concerns in purchasing decision.

## **Keywords**

Consumer's Purchasing Decision Analysis, Human Sensor, Product Exterior Design, Multivariate Analysis (MA)

## **1. Introduction**

In the product design, consumer preferences analysis is very helpful for consumer products design. One of the most important factors for a consumer's purchasing decision is a products exterior design. Products exterior design is dependent on sensibility of designer to demonstrate the features and functionalities of products. Hence, how to develop products exterior design which motivate consumers' affective responses on consumers' concerns is of great importance to an exterior design manufacturers. Traditionally, the success of a product's design dependent on the designer artistic sensibilities, which quite often did not meet with great acceptance in the marketplace. Generally, most design methodology researches address the importance of understanding customer needs during the product definition phase. Ullman (1992), and Otto and Wood (2001) provide sections about addressing customer needs. Bloch reported there are 5 constraints that always happened among designer and market desires followed ergonomic constraints; production and cost constraints; regulatory and legal constraints; marketing program constraints, and designer's constraints. In additional he said, a product's exterior design represent one solution to a set of design goals and constraints acted on by the designer and approved by management (Bloch, 2016).

In order to relate the product exterior design and costumer affective, the prediction models need to be constructed. Typically, two kinds of prediction models can be constructed, such as the classification-based model and the regression-based model. With the aid of such prediction model, an especially designed product exterior design that targets specific consumer groups can be produced more objectively and efficiently instead of only relying on the

designers' intuition and experience. Since various exterior design features would affect consumer's preference regarding product design. The issue of product exterior design aims to pin point critical product exterior design factors that influence customer concerns of the product design thus to aid designers in developing new products (Do. Bagus & Murata, 2016).

The MA was used to catching and analyze consumer perception toward a product exterior design. It is useful to build and describe the different segments of a market from a survey on potential consumers. When considering groups of objects in a multivariate data set, two situations can arise. Given a data set containing measurements on individuals, in some cases we want to see if some natural groups or classes of individuals exist, and in other cases, we want to classify the individuals according to a set of existing groups. Cluster analysis develops tools and methods concerning the former case, which is, given a data matrix containing multivariate measurements on a large number of individuals (or objects), the objective is to build some natural subgroups or clusters of individuals.

The MA framework will be proposed in section 2. Section 2 presents methodology used to analyze the factors that can attract consumer's purchasing decision for product exterior design, by looking ranked scores in decision table. Section 3 describes implementation procedures of MA and generate the model analyze for consumer's purchasing decision analysis. Section 4 presents the model evaluation by looking ranked scores of decision table. The conclusions and future work are presented in section 5.

## **2. Problem Statement**

Facing fierce competition in the marketplace, companies should try to determine the optimal settings of design attribute of new products from which the best customer's satisfaction can be obtained. Human perception becomes the main objective in product development. This is related to consumer taste and preferences. In order to realize a well-designed product that appeals to customers, the product not only should meet the physical requirements of consumers but also has to satisfy their affective needs. Thus, in a diverse and highly competitive market, successful products may not those with desirable product form.

Moreover, in design activities, such as involving the designer's decision, and so on at different levels, are a very stressful activity for everyone. No exception for the design engineer, the risks and consequences are very severe from the wrong decision on the viability of the company. However, the quality of a decision does not depend on the particular situation, but rather on the decision-making process used. Therefore, we need a decision-making support system for designers comprehensive.

In order to translate consumer desires to product attributes are needed a concept design from the product that had existed. The targeted products are evaluated from the consumers' viewpoint using affective words. By utilizing the data; such as a database, questionnaire and so on, the relationship between the affective word and the item or categories are decided. However, consumers obviously are not expert in product development. Thus, their sense related to the consumer affection are not precise. Thus, the researchers added the rough set model to increase the accuracy of the data from the consumers.

The fundamental concept behind Rough Set Theory is the approximation of lower and upper spaces of a set, the approximation of spaces being the formal classification of knowledge regarding the interest domain. The subset generated by lower approximations is characterized by objects that will definitely form part of an interest subset, whereas the upper approximation is characterized by objects that will possibly form part of an interest subset. Every subset defined through upper and lower approximation is known as Rough Set.

Knowledge classification is a fundamental problem in rough set theory. In Pawlak's rough set model, the degree of set overlap was not considered, namely, the classification must be totally correct or certain. Therefore, original rough set model cannot effectively deal with data sets which have noisy data and latent useful knowledge in the boundary region may not be fully captured. In order to overcome the limitations, some extended rough set models have been put forward which combine with other available soft computing technologies, such as statistical rough set model (Pawlak, Wong and Ziarko, 1988), decision-theoretic rough set model (Yao, et.al., 1992), fuzzy rough set model

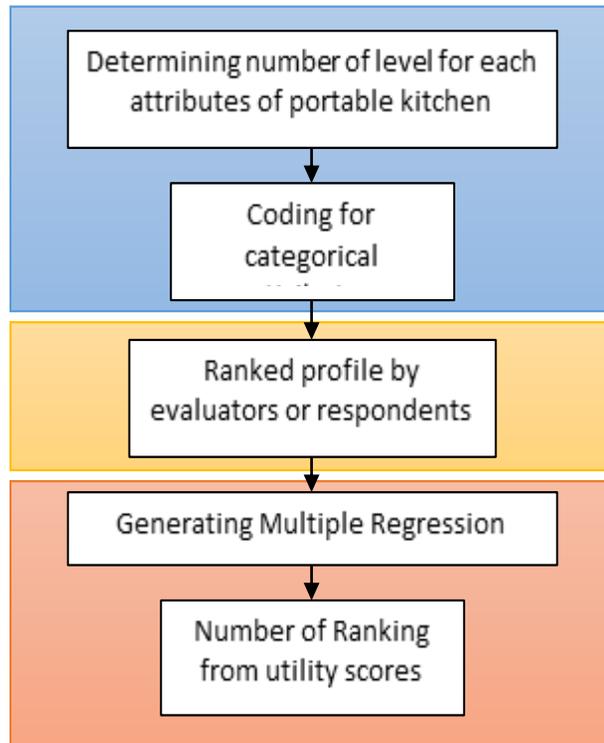


Figure 1. The Framework of Regression-based Model

(Dubois, et.al, 1992), covering rough set model (Zhu, et.al., 2003), tolerance rough set model (Skowron and Stepaniuk, 1996), dominance-based rough set model (Wojciech, et.al., 2008) and others.

This research aims to develop a methodology of the rough set with considering probability aspect. Since the rough set has several demerits which one of all is the inability to handle incomplete information in datasets. Knowledge classification is a fundamental problem in rough set theory. In Pawlak's rough set model, the degree of set overlap was not considered, namely, the classification must be totally correct or certain. Therefore, the original rough set model cannot effectively deal with data sets which have noisy data and latent useful knowledge in the boundary region may not be fully captured. The methodology gives several rules with higher confidence level than the original one, thus the variety of rule that extracted is more reliable even though the computational time taking a long time into processing.

### 3. Regression-based Model Structures

The framework of Regression-based model system is represented in "Fig. 1" which consisting of input, process, and output. The input consists of data for each attribute of portable kitchen and method, while the process is evaluation by respondents. The last is the output which consists of multiple regression equation and the result of ranking which the product attributes (PA) is more concerned by consumer as purchasing decision.

This paper conducts an empirical product study to investigate the relation between consumers' feelings (Human Sensory) in the market and product exterior design with portable kitchen design case study. Regression-based model is proposed to determine what the feature of product exterior design which attracts consumer's concern. This method is quite strength to solve vagueness, it's able to construct deterministic decision table using ranked score among evaluator.

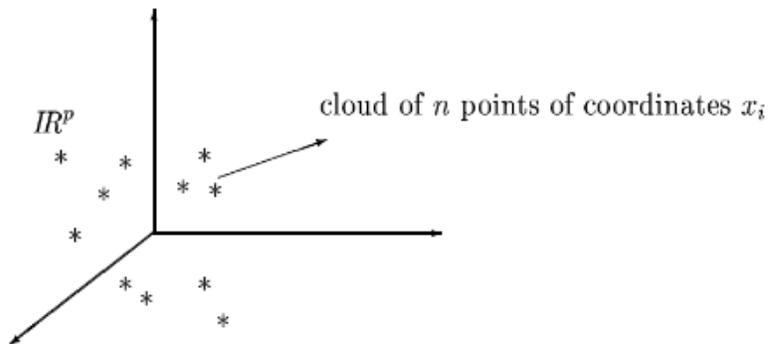
### 3.1 Determining level of portable kitchen for each attributes

Assume that the data matrix  $X(n \times p)$  is composed of  $n$  observations (or individuals) of  $p$  variables (level of attributes). There are in fact two ways of looking at  $X$ , row by row or column by column (Hardle, 2003):

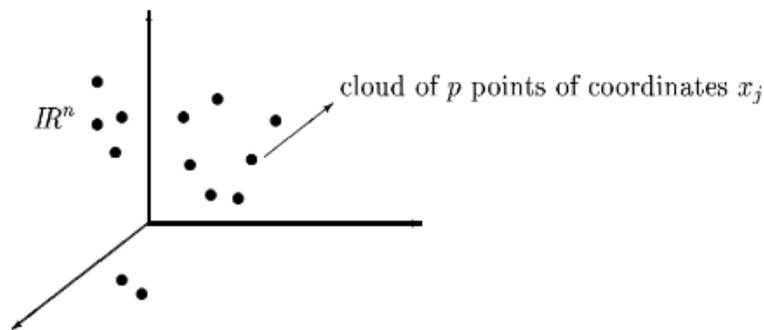
- (1) Each row (observation) is a vector  $x_i^T = (x_{i1}, \dots, x_{ip}) \in R^p$ .
- (2) Each column (variables) is a vector  $x_j = (x_{1j}, \dots, x_{nj})^T \in R^n$ .

When  $n$  and/or  $p$  are large (larger than 2 or 3), we cannot produce interpretable graphs of these clouds of points. Therefore, the aim of the factorial methods to be developed here is two-fold. We shall try to simultaneously approximate the column space  $C(X)$  and the row space  $C(X^T)$  with smaller subspaces, note that the space in this study is same meaning with level. The hope is of course that this can be done without losing too much information about the variation and structure of the point clouds in both spaces. Ideally, this will provide insights into the structure of  $X$  through graphs in  $R$ ,  $R^2$  or  $R^3$ . The main focus then is to find the dimension reducing factors. The different between row by row and column by column is depicted in "Fig. 2".

In this study,  $X$  is represented by a cloud of  $p$ -points (variables) in  $R^n$  (considering each column). The level of portable kitchen for each attributes is represented in "Table 1". We start as before with one dimension. In other words, we have to find a straight line  $G_j$ , which is defined by the unit vector  $v_j \in R^n$ , and which gives the best fit of the initial cloud of  $p$ -points. The representation of the  $j$ -th variable  $x_{[j]} \in R^n$  is obtained by the projection of the corresponding point onto the straight line  $G_j$  or the direction  $v_j$ . Hence we have to find  $v_j$  such that:



(a) The point of view for each row (observation)



(b) The point of view for each column (variables)

Fig. 2. The point of view in matrix  $X(n \times p)$ .

$$\max \sum_{j=1}^p \|p_{x_{[j]}}\|^2 \dots\dots\dots (1)$$

or equivalently, we have to find the unit vector  $v_1$  which:

$$\max(X^T v_1)^T (X v_1) = v_1^T (X X^T) v_1 \dots\dots\dots (2)$$

The coordinates of the p variables on  $G_1$  are given by:

$$u_1 = X^T v_1 \dots\dots\dots (3)$$

the first factorial axis. The  $p$  variables are now represented by a linear combination (see eq. 4) of the original individuals  $x_1, \dots, x_n$ , whose coefficients are given by the vector  $v_1$ , i.e., for  $j = 1, \dots, p$

$$u_{1j} = v_{11}x_{1j} + \dots + v_{1n}x_{nj} \dots\dots\dots (4)$$

### 3.2 Coding for Categorical

A frequently applied paradigm in analyzing data from multivariate observations is to model the relevant information (represented in a multivariate variable X) as coming from a limited number of latent factors. In a survey on household consumption, for example, the consumption levels, X, of p different goods during one month could be observed. The variations and covariations of the p components of X throughout the survey might in fact be explained by two or three main social behavior factors of the household. For instance, a basic desire of comfort or the willingness to achieve a certain social level or other social latent concepts might explain most of the consumption behavior. These unobserved factors are much more interesting to the social scientist than the observed quantitative measures (X) themselves, because they give a better understanding of the behavior of households.

This study represents factor analysis in order to categorize the attributes of portable kitchen into subspace or cloud of p-point of  $x_j$ . Factor analysis is a technique to identify factors that has relation of certain patterns inside the data sets. Like the original variables, the factors vary from individual to individual; but unlike the variables, the factors

Table 1. The level of portable kitchen for each attributes.

Attribute	Number of levels	Description of levels	Notation
Shape	4	Tube, string, oval, and flip-flap	X <sub>11</sub> , X <sub>12</sub> , X <sub>13</sub> , X <sub>14</sub> .
Storage capacity	3	Small, medium, large	X <sub>21</sub> , X <sub>22</sub> , X <sub>23</sub> .
Color	4	Monochrome, two colours, pattern, natural	X <sub>31</sub> , X <sub>32</sub> , X <sub>33</sub> , X <sub>34</sub> .
Material	4	Wood, plastic, metal, mixture	X <sub>41</sub> , X <sub>42</sub> , X <sub>43</sub> , X <sub>44</sub> .
Accessories	5	Slink, without accessories, wheel, cabinet for kitchen herbs, kitchen midden	X <sub>51</sub> , X <sub>52</sub> , X <sub>53</sub> , X <sub>54</sub> , X <sub>55</sub>

cannot be measured or observed. The existence of these hypothetical variables is therefore open to question. The aim of factor analysis is to explain the outcome of p variables in the data matrix X using fewer variables, the so-called factors. In practice, there are some data sets for which the factor analysis model does not provide a satisfactory fit. Thus, factor analysis remains somewhat subjective in many applications, and it is considered controversial by some

statisticians. Sometimes a few easily interpretable factors emerge, but for this data sets, neither the number of level nor the interpretation is clear. In order to overcome that issue, Do. Bagus, Muhammad Reza and Murata, Tomohiro (2016) proposed Conjoint Analysis into Kansei Engineering scheme. They said that the issue occurs caused from two characteristic of categorical attribute. Firstly, the PA vector is often sparse. Secondly, the difficulty of interpretation of the estimates for the recoded variables. However, they proposed the orthogonal coding to categorize the attribute which are shown to yield good predictions even when some profile combinations are not fully realistic. The predictions made from these designs are not subject to predictive bias if the correlation pattern among the attributes changes from the calibration set to the prediction set.

Table 2. The result of the system orthogonal scheme.

<i>Combination</i>	<i>Shape</i>	<i>Storage capacity</i>	<i>Color</i>	<i>Material</i>	<i>Accessories</i>
1	String	Medium	Natural	Wood	Wheel
2	String	Medium	Monochrome	Metal	Kitchen midden
3	Flip-flap	Large	Pattern	Plastic	Wheel
4	Oval	Medium	Monochrome	Plastic	Slink
5	Flip-flap	Small	Monochrome	Wood	Without accessories
6	String	Small	Two colours	Plastic	Cabinet
7	Flip-flap	Small	Natural	Mixture	Kitchen midden
8	Tube	Small	Two colours	Metal	Wheel
9	Tube	Large	Natural	Metal	Slink
10	Oval	Small	Natural	Wood	Cabinet
11	Tube	Large	Monochrome	Mixture	Cabinet
12	Oval	Large	Two colours	Wood	Kitchen midden
13	Flip-flap	Medium	Two colours	Wood	Slink
14	Tube	Medium	Natural	Plastic	Without accessories
15	Tube	Small	Monochrome	Wood	Wheel
16	Tube	Medium	Pattern	Wood	Kitchen midden
17	Flip-flap	Medium	Monochrome	Metal	Cabinet
18	Tube	Medium	Two colours	Mixture	Without accessories
19	String	Large	Monochrome	Wood	Without accessories
20	Oval	Small	Pattern	Metal	Without accessories
21	String	Small	Pattern	Mixture	Slink
22	Tube	Small	Monochrome	Plastic	Kitchen midden
23	Oval	Medium	Monochrome	Mixture	Wheel
24	Tube	Medium	Pattern	Wood	Cabinet
25	Tube	Small	Monochrome	Wood	Slink

Ideally all the information in X can be reproduced by a smaller number of factors. These factors are interpreted as latent (unobserved) common characteristics of the observed  $x \in R^p$ . The case just described occurs when every observed  $x = (x_1, \dots, x_p)^T$  can be written as:

$$x_j = \sum_{l=1}^k q_{jl}f_l + \mu_j, j = 1, \dots, p \dots\dots\dots (5)$$

Here  $f_l$ , for  $l = 1, \dots, k$  denotes the factors. The number of factors,  $k$ , should always be much smaller than  $p$ . For instance, in psychology  $x$  may represent  $p$  results of a test measuring intelligence scores. One common latent factor

explaining  $x \in R^p$  could be the overall level of “intelligence”. In marketing studies,  $x$  may consist of  $p$  answers to a survey on the levels of satisfaction of the customers. These  $p$  measures could be explained by common latent factors like the attraction level of the product or the image of the brand, and so on. Indeed it is possible to create a representation of the observations that is similar to the one in by means of principal components, but only if the last  $p - k$  eigenvalues corresponding to the covariance matrix are equal to zero. In practice the covariance matrix is rarely singular. It is common praxis in factor analysis to split the influences of the factors into common and specific ones. There are, for example, highly informative factors that are common to all of the components of  $X$  and factors that are specific to certain components. The factor analysis model used in praxis is a generalization of

$$X = QF + U + \mu \dots\dots\dots (6)$$

where  $Q$  is a  $(p \times k)$  matrix of the (*non-random*) loadings of the common factors  $F(k \times 1)$  and  $U$  is a  $(p \times 1)$  matrix of the (random) specific factors.

Table 3. The result of Pearson Correlation.

		Correlations																				
		Median	x11	x12	x13	x14	x21	x22	x23	x31	x32	x33	x34	x41	x42	x43	x44	x51	x52	x53	x54	x55
Pearson Correlation	Median	1.00	0.74	0.16	0.74	0.16	0.11	0.50	0.66	0.11	0.74	0.16	0.16	0.67	0.78	0.09	0.66	-0.37	0.02	-0.37	-0.02	0.08
	x11	0.74	1.00	-0.10	1.00	-0.10	0.04	0.60	0.91	0.04	0.80	-0.10	-0.10	0.58	0.53	0.15	0.72	-0.37	0.11	-0.19	-0.24	-0.14
	x12	0.16	-0.10	1.00	-0.10	1.00	0.75	0.09	-0.16	-0.27	-0.10	-0.22	-0.22	-0.17	0.38	0.38	0.02	0.01	-0.55	-0.16	0.02	-0.23
	x13	0.74	1.00	-0.10	1.00	-0.10	0.04	0.60	0.91	0.04	0.80	-0.10	-0.10	0.58	0.53	0.15	0.72	-0.37	0.11	-0.19	-0.24	-0.14
	x14	0.16	-0.10	1.00	-0.10	1.00	0.75	0.09	-0.16	-0.27	-0.10	-0.22	-0.22	-0.17	0.38	0.38	0.02	0.01	-0.55	-0.16	0.02	-0.23
	x21	0.11	0.04	0.75	0.04	0.75	1.00	-0.15	-0.04	0.00	-0.15	-0.10	-0.10	-0.20	0.32	0.32	-0.04	0.10	-0.46	-0.39	-0.21	-0.27
	x22	0.50	0.60	0.09	0.60	0.09	-0.15	1.00	0.53	-0.33	0.60	-0.28	-0.28	0.36	0.34	0.53	0.53	-0.19	-0.06	-0.01	0.15	0.07
	x23	0.66	0.91	-0.16	0.91	-0.16	-0.04	0.53	1.00	-0.04	0.91	-0.16	-0.16	0.51	0.45	0.08	0.82	-0.44	0.20	-0.09	-0.10	0.02
	x31	0.11	0.04	-0.27	0.04	-0.27	0.00	-0.33	-0.04	1.00	-0.15	0.92	0.92	0.20	-0.04	-0.39	-0.21	0.43	0.36	-0.23	-0.04	-0.08
	x32	0.74	0.80	-0.10	0.80	-0.10	-0.15	0.60	0.91	-0.15	1.00	-0.10	-0.10	0.58	0.53	0.15	0.91	-0.55	0.29	-0.01	-0.05	0.07
	x33	0.16	-0.10	-0.22	-0.10	-0.22	-0.10	-0.28	-0.16	0.92	-0.10	1.00	1.00	0.25	0.02	-0.34	-0.16	0.34	0.45	-0.16	0.02	-0.03
	x34	0.16	-0.10	-0.22	-0.10	-0.22	-0.10	-0.28	-0.16	0.92	-0.10	1.00	1.00	0.25	0.02	-0.34	-0.16	0.34	0.45	-0.16	0.02	-0.03
	x41	0.67	0.58	-0.17	0.58	-0.17	-0.20	0.36	0.51	0.20	0.58	0.25	0.25	1.00	0.51	-0.13	0.51	-0.24	0.32	-0.04	-0.13	-0.05
	x42	0.78	0.53	0.38	0.53	0.38	0.32	0.34	0.45	-0.04	0.53	0.02	0.02	0.51	1.00	-0.10	0.63	-0.44	-0.14	-0.44	-0.10	0.02
	x43	0.09	0.15	0.38	0.15	0.38	0.32	0.53	0.08	-0.39	0.15	-0.34	-0.34	-0.13	-0.10	1.00	0.08	0.26	-0.32	0.08	-0.10	-0.18
	x44	0.66	0.72	0.02	0.72	0.02	-0.04	0.53	0.82	-0.21	0.91	-0.16	-0.16	0.51	0.63	0.08	1.00	-0.61	0.20	-0.09	-0.10	0.02
	x51	-0.37	-0.37	0.01	-0.37	0.01	0.10	-0.19	-0.44	0.43	-0.55	0.34	0.34	-0.24	-0.44	0.26	-0.61	1.00	-0.21	0.19	-0.09	-0.12
	x52	0.02	0.11	-0.55	0.11	-0.55	-0.46	-0.06	0.20	0.36	0.29	0.45	0.45	0.32	-0.14	-0.32	0.20	-0.21	1.00	0.44	0.03	0.02
	x53	-0.37	-0.19	-0.16	-0.19	-0.16	-0.39	-0.01	-0.09	-0.23	-0.01	-0.16	-0.16	-0.04	-0.44	0.08	-0.09	0.19	0.44	1.00	-0.09	-0.12
	x54	-0.02	-0.24	0.02	-0.24	0.02	-0.21	0.15	-0.10	-0.04	-0.05	0.02	0.02	-0.13	-0.10	-0.10	-0.10	-0.09	0.03	-0.09	1.00	0.82
x55	0.08	-0.14	-0.23	-0.14	-0.23	-0.27	0.07	0.02	-0.08	0.07	-0.03	-0.03	-0.05	0.02	-0.18	0.02	-0.12	0.02	-0.12	0.82	1.00	

### 3.3 Multiple Regression Analysis

In this study, utilizing multiple linear regression one wants to determine the most appropriate linear model to predict only one dependent random variable  $y$  from a set of fixed, observed independent variables  $x_1, x_2, \dots, x_k$  measured without error. One aspect of interest will be choosing which variables to include in the model if this is not already known. Rencher, Alvin C stated that there were three cases in multivariate regression according to the number of variables [4]:

- (1) Simple linear regression: one  $y$  and one  $x$ . For example, suppose we wish to predict college grade point average (GPA) based on an applicant’s high school GPA.
- (2) Multiple linear regression: one  $y$  and several  $x$ ’s. We could attempt to improve our prediction of college GPA by using more than one independent variable, for example, high school GPA, standardized test scores (such as ACT or SAT), or rating of high school.

- (3) Multivariate multiple linear regression: several  $y$ 's and several  $x$ 's. In the preceding illustration, we may wish to predict several  $y$ 's (such as number of years of college the person will complete or GPA in the sciences, arts, and humanities). As another example, suppose the Air Force wishes to predict several measures of pilot efficiency. These response variables could be regressed against independent variables (such as math and science skills, reaction time, eyesight acuity, and manual dexterity).

To further distinguish case 2 from case 3, we could designate case 2 as univariate multiple regression because there is only one  $y$ . Thus in case 3, multivariate indicates that there are several  $y$ 's and multiple implies several  $x$ 's. The term multivariate regression usually refers to case 3.

There are two basic types of independent variables, fixed and random. In the preceding illustrations, all  $x$ 's are random variables and are therefore not under the control of the researcher. A person is chosen at random, and all the  $y$ 's and  $x$ 's are measured, or observed, for that person. In some experimental situations, the  $x$ 's are fixed, that is, under the control of the experimenter.

In the fixed- $x$  regression model, we express each  $y$  in a sample of  $n$  observations as a linear function of the  $x$ 's plus a random error,  $\varepsilon$ :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_q x_{1q} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_q x_{2q} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_q x_{nq} + \varepsilon_n \dots\dots\dots (6) \end{aligned}$$

The number of  $x$ 's is denoted by  $q$ . The  $\beta$ 's in eq. (6) are called regression coefficients. Additional assumptions that accompany the equations of the model are as follows:

- (1)  $E(\varepsilon_i) = 0$  ; for all  $i = 1, 2, \dots, n$ .
- (2)  $var(\varepsilon_i) = \sigma^2$  ; for all  $i = 1, 2, \dots, n$ .
- (3)  $cov(\varepsilon_i, \varepsilon_j) = 0$  ; for all  $i \neq j$ .

*Assumption 1* states that the model is linear and that no additional terms are needed to predict  $y$ ; all remaining variation in  $y$  is purely random and unpredictable. Thus if  $E(\varepsilon_i) = 0$  and the  $x$ 's are fixed, then  $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_q x_{iq}$ , and the mean of  $y$  is expressible in terms of these  $qx$ 's with no others needed. In *assumption 2*, the variance of each  $\varepsilon_i$  is the same, which also implies that  $var(y_i) = \sigma^2$ , since the  $x$ 's are fixed. *Assumption 3* imposes the condition that the error terms be uncorrelated, from which it follows that the  $y$ 's are also uncorrelated, that is,  $cov(y_i, y_j) = 0$ . Using matrix notation, the models for the  $n$  observations in eq. (6) can be written much more concisely in the form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1q} \\ 1 & x_{21} & \dots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \dots\dots\dots (7)$$

Or else the linear model can be written as:

$$y = X^* \beta^* + \varepsilon \dots\dots\dots (8)$$

The least squares solution is given by  $\hat{\beta}$ :

$$\begin{aligned} \hat{\beta} &= arg \min_{\beta} (y - X\beta)^T (y - X\beta) \\ &= arg \min_{\beta} \varepsilon^T \varepsilon \dots\dots\dots (9) \end{aligned}$$

Suppose that  $(X^T X)$  is of full rank and thus invertible. Minimizing the eq. (9) with respect to  $\beta$  yields:

$$\hat{\beta} = (X^T X)^{-1} X^T y \dots\dots\dots (10)$$

The coefficient of determination is influenced by the number of regressors. For a given sample size  $n$ , the  $r^2$  (coefficient determination) value will increase by adding more regressors into the linear model. The value of  $r^2$  may therefore be high even if possibly irrelevant regressors are included.

Table 4. The Results of Utility Scores.

<i>Attribute</i>	<i>Description of levels</i>	<i>Utility Scores</i>	<i>Rank</i>
Shape	Tube	-0.066	3
	String	-0.066	3
	Oval	0.131	2
	Flip-flap	0.154	1
Storage Capacity	Small	-0.336	3
	Medium	-0.202	2
	Large	0.368	1
Color	Monochrome	-0.066	3
	Two colours	-0.158	2
	Pattern	-0.066	3
	Natural	0.332	1
Material	Wood	0.085	3
	Plastic	0.347	1
	Metal	0.306	2
	Mixture	-0.348	4
Accessories	Slink	-0.347	5
	Without accessories	-0.205	4
	Wheel	-0.032	2
	Cabinet for kitchen herbs	-0.195	3
	Kitchen midden	0.189	1

#### 4. Application Regression-based Model to Portable Kitchen

The goal of this study is to decisive which the attributes of the portable kitchen that became consumer preferences as purchasing decision. The result was summarize into decision table with number of ranking. In order to implement the framework of MA, first of all, the PA and level of each PA were provided that shown in Table 1. The product attribute of portable kitchen design were consist of shape of body, storage capacity, color of body, material, and

accessories. Thus, by using orthogonal scheme into multivariate system the researcher generated 25 combination of product attribute that will be ranked by consumer advance. The combination is provided in Table 2.

Second of all, from the results of the ranking can be searched median and average rank to determine the data to be processed in a monotone transformation by using a computer program Statistical Analysis Software. Then levels on each combination is coded in the form of dummy variables with decision variables as follows:

$$y_n = \beta_0 + \dots + \beta_q x_{nq} \begin{cases} 1, \text{ if the PA is sufficient} \\ 0, \text{ Otherwise} \end{cases} \dots \dots \dots (11)$$

Third of all, the researcher got predictive model to calculate utility score. The highest utility score is indicated that the product attributes is more desired by consumer and it's also indicated as the rank of consumer purchasing decision toward a product design. The predictive model as followed:

$$y_n = 0.197x_{13} + 0.220x_{14} - 0.270x_{21} - 0.136x_{22} + 0.434x_{23} - 0.092x_{32} + 0.398x_{34} + 0.151x_{41} + 0.413x_{42} + 0.372x_{43} - 0.282x_{44} - 0.281x_{51} - 0.139x_{52} + 0.034x_{53} - 0.129x_{54} + 0.255x_{55} - 0.066 \dots \dots \dots (12)$$

Fourth of all, the researcher utilized utility function from the predictive model. Utility function as followed:

$$\hat{y}_n = \sum_{l=1}^{L_j} \beta_{jl} = 0 \dots \dots \dots (13)$$

$x_{jl}$  ( $l = 1, \dots, L_j$ ) are the levels of each factor  $X_j$  and the coefficients  $\beta_{jl}$  are the part-worths.

That is, one is interested in evaluating how well the model developed from the sample, often called the calibration, training, or model-building sample predicts future observations in a new sample called the validation sample. In model validation, one is investigating how well the parameter estimates obtained in the model development phase of the study may be used to predict a set of new observations. In multiple linear regression, the square of the population multiple correlation coefficient is used to measure the degree of linear relationship between the dependent variable and the population predicted value of the dependent variables,  $\beta'X$ . It represents the square of the maximum correlation between the dependent variable and the population analogue of  $\hat{y}$ . In some sense, the square of the multiple correlation coefficient is evaluating “model” precision. To evaluate predictive precision, one is interested in how well the parameter estimates developed from the calibration sample predict future observations, usually in a validation sample. One estimate of predictive precision in multiple linear regression is the squared zero-order Pearson product-moment correlation or *Pearson Correlation* between the fitted values obtained by using the estimates from the calibration sample with the observations in the validation sample.

Fifth of all, the researcher computed utility scores for each level of attributes that shown in Table 1 and give the ranking. The result of utility scores is provided in Table 4. Utility score was got from calculation of eq. (12) and Pearson Correlation (see Table 3).

The last step, the researcher computed the important factor of portable kitchen attribute that become consumer concern as purchasing decision by looking which of Table 3 have the highest ranking. For instance, the highest ranking of shape is flip-flap (it's shown with red-mark), so the flip-flap represent attribute of shape as consumer purchasing decision toward a product design of portable kitchen. The importance score of attribute is utilized eq. (14) as followed:

$$W_i = \frac{\beta_{max_{jl}} - \beta_{min_{jl}}}{\sum_{j=1}^k (\beta_{max_{jl}} - \beta_{min_{jl}})} \dots \dots \dots (14)$$

$\beta_{max_{jl}}$  is the highest level score from attributes  $j$ -th and level  $l$  and otherwise. While  $k$  is total attribute. The final result is ranking from product attributes that's provided in Table 4 and for the percentage of important factor for each attributes is provided in Fig. 3.

## **5. Discussion and Conclusion**

### **5.1 Discussion**

Twenty five combination product level, such as shown in Table 2, obtained from orthogonal process scheme of the level of portable kitchen for each attributes. Twenty five combination were evaluated by twenty five evaluator whose experiences in designer of product design, housewife and chef with random chooses.

Based on Table 3, the researcher found which of level of product attribute that become consumer purchasing decision toward a product design case, as followed:

- (1) For shape attribute, consumer were preferred “Flip-flap” and “Oval” as second choices. While “Tube” and “String” as third choices in shape attribute.
- (2) For storage capacity attribute, consumer were more desired a “Large” capacity than “Medium” and “Small” capacity.
- (3) For color attributes, consumer were preferred “Natural” color as the first choices, “Two colours” as the second choices, “Monochrome” and “Pattern” as the third choices.
- (4) For material attribute, consumer were desired “Plastic” as the first choices, “Metal” as the second choices, “Wood” and “Mixture” as the third and fourth choices.
- (5) For accessories attribute, the first chooses is “Kitchen midden”, “wheel” as the second choices, “Cabinet” as the third choices, “without accessories” as the fourth choices, and “slink” as the fifth choices.

Furthermore, based on Table 4, the researcher got the ranking for each attributes as followed:

- (1) The first rank, “Storage capacity” is more considered for consumer’s as purchasing decision.
- (2) The second rank, “Material” is considered lower than “Storage capacity” for consumer’s as purchasing decision.
- (3) The third rank, “Accessories” is more considered as low priority under the “Storage capacity” and “Material”.
- (4) The fourth rank, “Material” is considered as low priority under “Storage capacity”, “Material” and “Accessories” respectively.
- (5) The last rank, “Shape” is considered not important from all attribute that evaluated.

### **5.2 Conclusion**

Regression-based model is promising technique in consumer-oriented technology and quite effectively extracted costumer perception on the product design. This is also strengthened from the result of this study. For future work, there several point that should consider are:

- (1) Test the alignment of the model estimates should be evaluated, if the dummy variables used in the regression, the value of  $R^2$  will show the extent to which the model fits the data.
- (2) Test the back of reliability can be done by duplicating most of the analysis results with the retest way to some other respondents to see the consistency of the results.
- (3) To test validation can be predicted based on the function of most who have estimated, then evaluate the predicted correlated with results obtained from research in the field.

### **Acknowledgements**

We would also like to show our gratitude to the Professor Tomohiro Murata from Graduate School of Information, Production, and System, Waseda University, for sharing his pearls of wisdom with us during the course of this research. We are also immensely grateful to Muhammad Sofyan D. M., Director of STMIK Tidore Mandiri for his comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

## References

- B. Rundh, "Packaging design: creating competitive advantage with product packaging," *Br. Food J.*, vol. 111, no. 9, pp. 988–1002, 2009.
- Cameron, C., Trivedi, P., *Microeconometrics: Methods and Applications*. Cambridge University Press, New York, 2005.
- Do. Bagus, M. Reza, and Murata, T., Conjoint Analysis of Costumers' Preferences with Kansei Engineering System for Product Exterior Design, *Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics*, pp 1026-1032, Kumamoto, Japan, 2016.
- G. Cestre and R. Y. Darmon, "Assessing consumer preferences in the context of new product diffusion," *Int. J. Res. Mark.*, vol. 15, no. 2, pp. 123–135, 1998.
- H. R. Moskowitz, M. Reisner, J. Ben Lawlor, and R. Deliza, "Patterns in Packages: Learning from Many Packages and Many Attributes," *Packag. Res. food Prod. Des. Dev.*, pp. 35–41, 2009.
- H. Zhang, J. Zhou, D. Miao, and C. Gao, "Bayesian rough set model: A further investigation," *Int. J. Approx. Reason.*, vol. 53, no. 4, pp. 541–557, 2012.
- K. Marsh and B. Bugusu, "Food packaging - Roles, materials, and environmental issues: Scientific status summary," *J. Food Sci.*, vol. 72, no. 3, 2007.
- P. H. Bloch and P. H. Bloch, "Seeking the Ideal Form : Product Design and Consumer Response Published by : American Marketing Association Stable URL : <http://www.jstor.org/stable/1252116> REFERENCES Linked references are available on JSTOR for this article : You may need to log in to J," vol. 59, no. 3, pp. 16–29, 2016.
- S. Schütte, "Evaluation of the affective coherence of the exterior and interior of chocolate snacks," *Food Qual. Prefer.*, vol. 29, no. 1, pp. 16–24, 2013.
- T. Nishino, M. Nagamachi, and H. Tanaka, "Variable Precision Bayesian Rough Set Model and Its Application to Kansei Engineering," *Trans. Rough Sets V*, pp. 190–206, 2006.
- W. Hardle, "Applied Multivariate Statistical Analysis," no. April, 2003.
- Washington, S., Karlaftis, M., Mannering, F., *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC Press, Boca Raton, FL. 2003.
- Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. ...*, pp. 1–51, 1982.

## Biographies

**Muhammad Reza Do. Bagus** currently held Master of Engineering degree from Graduate School of Information, Production and System at Waseda University, Japan. Mr. Do. Bagus also held Bachelor of Engineering degree in Industrial Engineering field. This study reflects his thesis's work which is focused on developing a new Simulation-Optimization approach in data mining. Currently, he is a Lecturer in Information System Department at School of Information Management and Computer Tidore Mandiri, and his research interests include simulation, simulation optimization, data mining, machine learning, and supply chain management. He also represent his campus in The 5th IIAE International Conference on Industrial Application Engineering 2017 and his paper is chosen as "Best Student Paper Award" in that conference.