Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 1, No. 4, pp. 89-95, 2023

Determining Customer Preferences in Choosing a Marketplace Using the Conjoint Analysis Method

Sri Agustina[1*], Haposan Sirait[2], Yasir Salih[3]

[1,2]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Univeristy of Riau, Indonesia*
[3]*Department of Mathematics, Faculty of Education, Red Sea University, Sudan*

**Corresponding author email: sri.agustina1864@student.unri.ac.id*

Abstract

The number of online shopping transactions in Indonesia has grown over the last ten years by 17% and the total number of e-commerce businesses has reached 26.2 million units. This creates competition that requires companies to maintain their existence by understanding consumer psychology. This study aims to determine the combination of attribute levels that are most preferred by consumers in choosing a marketplace as a place to shop by using conjoint analysis. The data used is the result of a survey of Riau University students in the form of a questionnaire. The results of the conjoint analysis in this study show that the level of each attribute that respondents prefer is the display of full color applications and images, free shipping promotions on a certain amount of purchase, the method of paying Cash On Delivery (COD), using J&T/JNE/Sicepat/Tiki/Pos delivery services, and product reviews are available in the form of photos and videos.

Keywords: Marketplace, Conjoint analysis, Customer preferences.

1. Introduction

According to Harahap & Amanah (2018) that marketplace or e-commerce is a transaction process carried out through media or intermediaries, which can be in the form of sites that provide goods and services or through social media. There are many marketplaces that can be accessed in Indonesia, both international and national marketplaces, and even local product marketplaces where more and more regional startup marketplaces are also developing.

Purchases of products via e-commerce in Indonesia reached IDR 146.7 trillion, an increase of 41% from IDR 74 trillion in 2015. The increasing growth of e-commerce in Indonesia has created intense competition, requiring companies to maintain their existence. The way that can be done is to understand consumer psychology, namely perceptions, motivations, and attitudes that drive purchasing decisions. Marketplace companies are trying to gain a larger market share than competitors, therefore marketplaces must consider the basis for consumer purchasing decisions (Viora & Suyanto, 2020). A marketplace needs to know something that is on the minds of consumers in determining segmentation and a decision-making process in purchasing. The next stage is knowing consumer responses regarding product choices, quantity and timing of purchases, and brands (Fahimah & Muyassaroh, 2020).

Conjoint analysis is used to determine a person's perceptions and preferences of an object. Conjoint analysis will provide a quantitative measure of the level of utility and relative importance of an attribute compared to other attributes (Widyawati et al., 2014). Previous research on conjoint analysis has been carried out by (Wulandari et al., 2019), (Wijayanto et al., 2007), and many other researchers.

2. Materials and Methods

2.1. Materials

This article was conducted using TPT, KP, PE, TPAK, and UMR data on Sumatra Island in 2014-2019 obtained from BPS Riau Province.

2.2. Methods

According to Pedigo & Buntin (1994) that some of the advantages of sampling include reduced costs, greater speed, greater coverage and greater level of accuracy. Quota sampling is a technique for determining samples/respondents up to the required quota. The characteristics of the respondents and the number of quotas taken were based on the considerations of the researcher.

According to Lord & Novick (2008) validity indicates the extent to which a measuring device can measure a criterion. The validity of the instrument was tested using the correlation coefficient between the scores of the question items and the total score calculated using the correlation coefficient formula. According to Omar et al. (2012) correlation coefficient can be calculated using the formula:

$$r_{hitung} = \frac{N\sum_{i=1}^{n} P_i T - (\sum_{i=1}^{n} P_i)(\sum T)}{\sqrt{(N\sum_{i=1}^{n} {P_i}^2 - (\sum_{i=1}^{n} P_i)^2)(N\sum T^2 - (\sum T)^2)}} \quad (1)$$

By r_{hitung} stating the correlation between the score of the question items and the total score of the questions, Nthe number of respondents, the score of the P_i *1st* instrument , from $i = 1,2,3 \dots, n$.

The validity test hypothesis is:

H_0 : There is no correlation between questionnaire questions.

H_1 : There is a correlation between the questionnaire questions.

Reliability in measurement implies that the measure or instrument has stability and consistency in it. (Pedhazur & Schmelkin, 1991) . The thing that is used as a reference in the reliability test is Cronbach's alpha , the formula is:

$$R = \left(\frac{l}{l-1}\right)\left(1 - \frac{\sum_{b=1}^{k} \sigma_b^2}{\sigma_t^2}\right) \quad (2)$$

with, R stating Cronbach's Alpha coefficient, l the number of questions, $\sum_{b=1}^{l} \sigma_b^2$ the total variance of the question items, where $\sigma_b^2 = \frac{\sum(P_i^2 - (\frac{\sum P_i}{N})^2}{N}$, $\sum_{b=1}^{l} \sigma_b^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2$, σ_t^2 represents the total variance where $\sigma_t^2 = \frac{\sum(P_t^2 - (\frac{\sum P_t}{N})^2}{N}$.

The reliability test hypothesis is:

H_0 : No reliability on the instrument

H_1 : There is reliability in the instrument

Hypothesis rejection criterion is reject H_0if the Alpha coefficient is greater than the significance level of 70% or 0.7 (Strainer, 2003).

According to Montgomery (2017) that in factorial designs with 2^k larger factor values k, the number of complete combinations will also be even greater, this will complicate the resources of most researchers. Fractional factorial design with two levels is denoted by 2^{k-p}. This means that from the complete combination design 2^k, only the combination design 2^{k-p} was used for the experiment.

The general equation for multiple linear regression can be written as below (Montgomery, 2017).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (3)$$

with, β_0 denotes a constant or intercept, β_j the regression parameter, X_j denotes the independent variable, from $j = 1,2, \dots, k$. Based on equation (3) it can be broken down as a whole according to the scalar method to:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1, \\ Y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2, \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \varepsilon_n \end{aligned} \quad (4)$$

Equation (5) can be written in matrix notation as below:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & & \vdots\ \vdots\ \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Estimation of the regression parameters using the least squares method is carried out by first determining the sum of the squared errors. Here are the similarities:

$$\mathbf{S} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{5}$$

then minimize the sum of the squared errors , to get the least squares estimator $\widehat{\boldsymbol{\beta}}$ which minimizes $\boldsymbol{S}$ required that $\frac{\partial \mathbf{S}}{\partial \boldsymbol{\beta}|\boldsymbol{\beta}}$ obtained then:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{6}$$

According to Gujarati & Porter (2010) that in regression analysis if the independent variable is a nominal scale, then it is called regression with a dummy variable. The form of the regression equation with dummy variables is:

$$Y = \beta_0 + \beta_1 D_{1j} + \beta_2 D_{2j} + \cdots \beta_k D_{kj} + \varepsilon \tag{7}$$

with, D_{ij} denotes a dummy variable, of $i = 1,2, \ldots, k, j = 1$ if the corresponding value occurs, another 0.

According to Wirth (2014) that conjoint analysis is a multivariate technique that is usually used to determine the usefulness value of multidimensional product features and is very suitable for measuring human perceptions and preferences.

According to Hair et al. (2014) there are several provisions in choosing the method to be used in designing a combination of levels of attributes based on the number of attributes used, namely:

a. Choice-Based Conjoint (CBC)
   Choice-Based Conjoint (CBC) can be used if the number of attributes is less than or equal to six.
b. Traditional Conjoints
   Traditional Conjoint can be used with a maximum of ten attributes.
c. Adaptive Conjoint Analysis (ACA)
   Adaptive Conjoint Analysis (ACA) can be used with a maximum number of thirty attributes.

The approach used in designing attribute combinations is:

a. Pairwise combination
   Pairwise combination is an evaluation of two assessment factors. Respondents rated the two attributes until all possible pairs of the two attributes had been evaluated.
b. Complete combination (full profile combination)
   This method is used to evaluate stimuli formed from all attributes.

Conjoint analysis can be done with several procedures as follows.

a. Determine the basic model of conjoint analysis
   The basic model of conjoint analysis is as follows:

$$U(x) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} a_{ij} x_{ij} \tag{8}$$

   with,
   $U(x)$: The overall utility or NKL of each level in each attribute,
   a_{ij} : Level Usability Value (NKL) to- on the ji-*th* attribute
   x_{ij} : The *dummy* level variable to-j attribute to-i
   The method that can be used to complete the model from conjoint analysis is the least squares method with dummy variables.
b. Calculating Level Usability Value
   Calculating the Level Usefulness Value or the utility value of each level with two attribute levels using the following formula:

$$a_{ij} - a_{ik_i} = \beta_{ij} \tag{9}$$

$$a_{ij} + a_{ik_i} = 0 \tag{10}$$

c. Calculating Important Relative Values
   Important Relative Value (NRP) aims to determine the level of relative importance of an attribute to other attributes. Important Relative Value is calculated using the formula:

$$W_i = \frac{I_i}{\sum_i^m I_i} \tag{11}$$

so that $\sum_i^m W_i = 1$, where $I_i = \{\max(a_{ij}) - \min(a_{ij})\}$ for each i. By, W_i stating the Relative Important Value (NRP) of the attribute i, I_i attribute importance value for the attribute i.

According to Draper & Smith (1998), the simultaneous test is used to determine the relationship between the independent variable and the dependent variable simultaneously (simultaneously). This test will be carried out using the F test. The hypotheses that are tested simultaneously are:

H_0 : There is no influence between the independent variables on the dependent variable $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1: There is influence between the independent variables on the dependent variable $\beta_j \neq 0,$ for j

The test statistics used are:

$$F_{hitung} = \frac{MS_R}{MS_{Res}} \tag{12}$$

Decision : Reject H_0 if F_{hitung} greater than F_{tabel} $(F_{\alpha;(k-1,n-k)})$.

Partial test is used to see the relationship of independent variables to the dependent variable individually (Draper & Smith, 1998). The test statistic used in the partial test is the student's t-test. The partial test hypothesis used is as follows:

H_0: There is no effect of the independent variable $k$on the dependent variable $\beta_k = 0$

H_1: There is the effect of the independent variable $k$on the dependent variable $\beta_k \neq 0$, $k = 1,2, \dots, p$

Test statistics used:

$$t = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \tag{13}$$

where

$$\hat{\beta}_k = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$se(\hat{\beta}_k) = \sqrt{\frac{\sum_{i=1}^n y_i^2 - \hat{\beta}_k \sum_{i=1}^n y_i x_i}{n-1-k}}$$

Decision: Reject H_0 if $|t_{hitung}|$ greater than t_{tabel} $(t_{\alpha/2;n-k)})$.

3. Results and Discussion

The validity test was carried out to see whether or not there was a correlation between questions indicating that the questionnaire was declared valid or not. The questionnaire is declared valid if there is a correlation between questions. The data used to test the validity of 100 respondents. Validity test using the correlation formula r. The validity test hypothesis is:

H_0 : There is no correlation between questionnaire questions.

H_1 : There is a correlation between the questionnaire questions.

The confidence level used is 95% ($\alpha = 0.05$). The rejection criterion is H_0 if $r_{hitung} \geq r_{tabel}$. Based on the test results, the validity test results are obtained as Table 1.

Table 1: Instrument Validity Test Results

Combination to	r_{hitung}	r_{tabel}	Conclusion
1	0.5864539	0.195	valid
2	0.6772293		valid
3	0.7047701		valid
4	0.8181678		valid
5	0.7715939		valid
6	0.6946926		valid
7	0.7575313		valid
8	0.6919283		valid

From table 1 it can be seen that, in the first combination to the 8th combination the values $r_{hitung} \geq r_{tabel}$ the decision is reject H_0, meaning that there is a correlation in each combination. In conclusion, combinations 1 to 8 are valid.

The reliability test was carried out to see the stability and consistency of the research instrument. The data used in this reliability test were 100 respondents. The reliability test was performed using Cronbach's alpha formula. The reliability test hypothesis is as follows:

H_0 : There is no reliability of the questionnaire

H_1 : There is reliability of the questionnaire

Using a 95% confidence level ($\alpha = 0.05$), the rejection criterion is reject H_0 if $r_{hitung} \geq 0.7$.

From the results of calculations using Rstudio, the Cronbach alpha value is 0.86. due to Cronbach's alpha value ≥ 0.7, then reject H_0, meaning that the questionnaire is reliable, or can provide consistent results as a survey measuring tool.

Conjoint model estimation is done by coding using dummy variables. The following is a dummy variable that represents each attribute in this study in the Table 2.

Table 2: Coding of Dummy Variables for Each Attribute

Attribute	Levels	*Dummy* variable
Application view	Simple display	1
	Full color display and images	0
Promo	Product discount on a certain purchase amount	1
	Free shipping on a certain purchase amount	0
Payment method	Bank transfers/Indomaret/Alfamart	1
	Pay on the spot (COD)	0
Delivery service	J&T/JNE/Sicepat/Tiki/Pos	1
	The delivery service belongs to *the marketplace*	0
Review	Product reviews are available in the form of comments and ratings	1
	Product reviews are available in the form of photos and videos	0

Then an experiment was carried out using a fractional design to get a combination of stimuli as shown in Table 3 below.

Table 3: Dummy variables for each stimulus

Stimuli	X_1	X_2	X_3	X_4	X_5
1	1	1	1	1	1
2	0	0	0	1	1
3	1	0	1	0	1
4	0	1	0	0	1
5	0	1	1	1	0
6	1	0	0	1	0
7	0	0	1	0	0
8	1	1	0	0	0

Estimation of the regression model was carried out using the least squares method, the following regression models were obtained:

$$Y = 7.3725 - 0.2825X_1 - 0.2275X_2 - 0.1800X_3 + 0.5825X_4 - 0.2125X_5$$

The regression model parameter test was carried out using a partial test and a simultaneous test. Partial test obtained using the help of Rstudio. The partial test hypothesis is:

H_0 : There is no significant influence of variables

H_1 : There is a significant influence of variables

Following are the results of the partial test of the regression model shown in Table 4 below.

Table 4: Partial Test Results

Variable	$\lvert t_{hitung} \rvert$	t_{tabel}	Conclusion
X_1	3.357		Significant
X_2	2.703		Significant
X_3	2.139	1.96145	Significant
X_4	6.921		Significant
X_5	2.525		Significant

The significance level used $\alpha = 0.05$, because the value $|t_{hitung}|$ of the finite variable X_1 is X_5 greater than t_{tabel}, then all variables are rejected H_0. In conclusion, there is a significant effect of the variable on the X_1 finite variable X_5.

Simultaneous test is used to see the suitability of the regression model obtained. The hypothesis is:
H_0 : There is no model match
H_1 : There is model compatibility
Significance level $\alpha = 0.05$, the rejection criterion is reject H_0 if it F_{hitung} is greater than F_{tabel}.

Table 5: Simultaneous Anova Test

Source of variation	Squared sum	Free degrees	Average squared	F_{hitung}
Regression	219.370	5	43.874	15.485
Error	4516.380	1594	2.833	
Total	4735.750	1599		

Simultaneous tests carried out using the help of Rstudio obtained F_{hitung} is 15.48 and F_{tabel} is 2.219712 because $F_{hitung} > F_{tabel}$, then reject H_0. In conclusion, there is a suitability of the regression model.

Conjoint analysis to determine customer preferences in choosing a marketplace as a place to shop online can be seen from the Level Usability Value of each attribute.

The application view attribute has two levels namely the simple view, and the full color and image view. Level Usability Value on full-color displays and images has a positive NKL value of 0.1412 while a simple display has an NKL of -0.1412. This means that the full color and image view of the application is preferred by customers compared to the simple display of the application display attributes.

The promo attribute has levels, namely product discounts for a certain purchase amount and free shipping for a certain purchase amount. The free shipping level for a certain purchase amount has a positive NKL of 0.1138 while the product discount level for a certain purchase amount has a negative NKL of -0.1138. This means that in the promo attribute, customers prefer free shipping on a certain purchase amount compared to a discount on a certain purchase amount.

The payment method attribute has levels, namely bank transfers/Indomaret/Alfamart and cash on delivery (COD). The level of pay on the spot (COD) has a positive NKL of 0.0900 while bank/Indomaret/Alfamart transfers have a negative NKL of -0.0900. This means that customers prefer to pay on the spot (COD) compared to bank/Indomaret/Alfamart transfers on the payment method attribute.

The delivery service attribute has levels, namely J&T/JNE/SICEPAT/TIKI/POS and the delivery service belongs to the marketplace . The J&T/JNE/SICEPAT/TIKI/POS level has a positive NKL of 0.2912 while the delivery service level belonging to the marketplace has a negative NKL of -0.2912. This means that the J&T/JNE/SICEPAT/TIKI/POS level is preferred by customers compared to the delivery service level belonging to the marketplace on the delivery service attribute.

The review attribute has a level of available product reviews in the form of comments and ratings, another level is available product reviews in the form of photos and videos. The available level of product reviews in the form of photos and videos has a positive NKL of 0.1063, while available product reviews in the form of comments and ratings have a negative NKL of -0.1063. This means that the level of available product reviews in the form of photos and videos is preferred by customers compared to available product reviews in the form of comments and ratings on the review attribute.

Important Relative Value is used to determine the level of importance of the attribute from other attributes. Important Relative Value can be seen in the Table 6 below.

Table 6: Important Relative Value

Attribute	NRP	NRP x 100%
Application view	0.1902	19.0168
Promo	0.1533	15.3266
Payment method	0.1212	12.1212
Delivery service	0.3922	39.2189
Review	0.1432	14.3165

From the table it can be seen that the delivery service attribute has the highest relative importance value, namely 0.3922 or about 39% of all attributes, followed by application appearance, then promotions, reviews and finally the payment method has the lowest relative importance value, namely 0.1212 or around 12%.

4. Conclusion

From the results of this study, it can be concluded that the most influential attribute is the delivery service. The level that respondents prefer for each attribute is a full color and image application display, free shipping promo on a certain purchase amount, payment on delivery (COD) method, using J&T/JNE/Sicepat/Tiki/Pos delivery services, and available product reviews in the form of photos and videos.

References

Draper, NR, & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley & Sons.

Fahimah, M., & Muyassaroh, EM (2020). Buying interest on unicorn marketplaces in Indonesia: Tokopedia and Bukalapak. *Journal of Business & Accounting* , *9* (1), 31–42.

Gujarati, DN, & Porter, DC (2010). *Basics of econometrics* . Jakarta: Salemba Empat.

Hair, JF, Black, WC, Babin, BJ, & Anderson, RE (2014). *Multivariate data analysis* (7th ed.). United States Of America: United States Of America: Pearson Education Limited.

Harahap, DA, & Amanah, D. (2018). Online Shopping Behavior in Indonesia: Case Study. *JRMSI - Indonesian Science Management Research Journal* , *9* (2), 193–213.

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Montgomery, DC (2017). *Design and analysis of experiments* . Arizona: John Wiley & Sons Inc.

Omar, A., Reusch, U., Knackmuss, S., Little, M., & Weiss, S. F. (2012). Anti-LRP/LR-specific antibody IgG1-iS18 significantly reduces adhesion and invasion of metastatic lung, cervix, colon and prostate cancer cells. *Journal of molecular biology*, *419*(1-2), 102-109.

Pedhazur, EJ, & Schmelkin, LP (1991). *Measurement, design, and analysis: An integrated approach* . New Jersey: Lawrence Erlbaum Associates.

Pedigo, L. P., & Buntin, G. D. (Eds.). (1994). *Handbook of sampling methods for arthropods in agriculture* (No. 04; SB933. 14, P4.). Boca Raton, FL: CRC press.

Strainers, DL (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment* , *80* (1), 99–103.

Viora, P., & Suyanto, A. (2020). Purchase decisions based on perceptions, motivations, and attitudes to the marketplace. *Journal of Business and Management Research* , *13* (1), 16–22.

Widyawati, W., Sitepu, R., & Napitupulu, N. (2014). Application of conjoint analysis of student preferences for work (case study: USU FMIPA undergraduate students). *Saintia Mathematics* , *2* (2), 189–200.

Wijayanto, H., Angraeni, Y., & Riskinandini, R. (2007). Conjoint analysis: full profile and cbc methods to examine student perceptions of job choices. *Statistics and Computing Forum* , *12* (1), 8–17.

Wirth, FF (2014). Consumers' shrimp purchasing preferences: an application of conjoint analysis. *Journal of Food Products Marketing* , *20* (2), 182–195.

Wulandari, S., Sari, FM, & Padang, UN (2019). The preference of the Padang city community for the use of online motorcycle taxi services in 2019. *Statistics* , *7* (2), 124–132.