

## CHAPTER III

### RESEARCH METHODOLOGY

#### A. Research Subject

The subject of this research are the respondents of PDAM Tirta Marta household customers in Yogyakarta City who are still active without distinguishing the customer tariff class.

#### B. Types of the Data

The data used in this research is the primary data and secondary data. According to Hasan (2002) the primary data becomes the data faster or more directly in the field by the person doing the research or the one who needs it. Primary data can be from sources of informants or individuals such as the results of interviews conducted by researchers. Primary data were obtained from interviews with parties related information as well as interviews with the respondents through the questionnaire where respondents are PDAM Tirtamarta household customers in Yogyakarta City. Primary data in this research are family income, education, house ownership, the family members, duration of stay, water quality, and customer satisfaction.

Secondary data is data obtained or collected by people conducting research from existing sources (Hasan, 2002). This data is used to support the primary information that has been obtained from the library materials, literature, previous research, books, and so on. Secondary data in this

research conducted of PDAM Tirtamarta Yogyakarta, Environment Agency of Yogyakarta, Central Bureau of Statistics.

### **C. Sampling Method**

Population refers to all groups of people, events, or things that are important to be meticulous (Sekaran and Bougie, 2013). Population in this research was taken from PDAM Tirtamarta household customers in Yogyakarta City. The methods used in this research is accidental sampling, Accidental sampling is a sampling technique that someone incidentally found and can be used as a sample when the requirement which is household customer. Accidental sampling is taking the sample is accidental by taking the case or the respondent who happens to be present or available somewhere in accordance with research context (Notoatmodjo, 2010). Based on the samples, respondents are anyone who easily found by investigators but still in the scope of a requirement, which are household customers of PDAM Tirtamarta in Yogyakarta City.

From the population then determined the number of samples, The sample covers part of the population and is considered representative of the population (Sekaran and Bougie, 2010), In the sample size determination of the author using the formula developed by Isaac and Michael took from Sugiyono (2011) in Quantitative Research Methods, Qualitative and R & D,

in this research using Theory Isaac and Michael for a high error of 10% with the formula:

$$S = \frac{\lambda^2 \cdot N \cdot P \cdot Q}{d^2 (N-1) + \lambda^2 \cdot P \cdot Q}$$

$$S = \frac{0,10^2 \cdot 29.954 \cdot 0,5 \cdot 0,5}{0,5^2 (29.954-1) + 0,10^2 \cdot 0,5 \cdot 0,5}$$

Description:

$\lambda^2$  : where the  $\lambda^2$  level is determined by dk equal to 1, the error rate of 10%

$P = Q = 0,5$

$D = 0,5$

$N$  = Number of samples of household customers of PDAM Tirtamarta 2016

$s$  = number of samples

After calculated used Isaac and Michael Formula, from the total household customers that are 29.954 then the total of samples obtained amounted to 269 respondents. In this research the total samples was made into 308 respondents by adding error rate 10%.

#### **D. Data Collection Method**

Research methods used in this research is two technique for collection data those are:

1. Questionnaire, The questionnaire is a data collection technique by providing a list of statements and questions that the author will ask the

respondent which means is a technique of collecting information through a sheet containing a list of questions to obtain information from a number of respondents.

2. Research in Librarianship, is the form of activities to research a number of literature such as books, internet, journals, articles and other mass media related to the problem under research to obtain information related to research.

### **E. Definition of Operational Variable**

The variables used in this research are dependent variable and independent variables, the variables used are as follows:

1. Dependent Variable,

Willingness to pay in this research is for quality improvement of household customers of PDAM Tirtamarta in Yogyakarta City that are expressed in units of the Indonesian Rupiah (IDR).

In the WTP calculated how far the ability of any individual or the public in issuing money in order to pay for quality improvement of PDAM Tirtamarta against the replacement of leaky pipes with new pipes to reduce water loss in Yogyakarta City. WTP is one part of the CVM method (Contingent Valuation Method) that will be used in this research.

Contingent valuation method is a direct method of economic valuation through willingness to pay (WTP) by someone, in this research using dichotomous CVM, such as Yes/No and others. The

calculation of the value of the WTP of CVM can do with logistics, probit and turnbull methods. This research used a logistics method (binary logistics method) to get value of the WTP of society for quality improvement of PDAM Tirtamarta in Yogyakarta City.

The willingness to pay (WTP) estimation is obtained through the average value of the willingness to pay (EWTP) of the respondents in the Focus Group Discussion (FGD) with Bidding Game method which will have the overall sum of Willingness to pay and divided by the number of respondents. The estimated average willingness to pay values defined as follows:

$$EWTP = \frac{\sum_{i=1}^n WTP_i}{n}$$

Which are:

*EWTP* : Estimated average of WTP

*WTP<sub>i</sub>* : Value of WTP to-i

*n* : Number of respondents

*I* : Respondent –i who are willing to pay (1,2,3,...n)

This EWTP value is used to determine the willingness to pay respondent with Dichotomous Choice. The dummy variable value of WTP is 1 if "WTP = EWTP" and 0 if "WTP ≠ EWTP" Independent Variables

2. The independent variable is the variable affecting the dependent variable, either positively or negatively (Sekaran and Bougie, 2010), independent variables in this research are as follows:

- a. Family Income (INC) is the wage or salary of the respondent obtained each month, for the housewife its revenue is the totally monthly consumption expenditure in the form of the Rupiah (IDR). This variable refers to the monthly household income. This includes income from all sources that are still living in one household. By looking at the amount of income then the respondent will take the decision of willingness to pay or not. How the amount of income earned by the respondent affect the willingness of consumers to pay for water quality improvement of PDAM Tirtamarta (measured in IDR).
- b. Education Level (EDU) shows formal education already or being taken by someone. The higher a person's level of education then the thought of insight and his views will be wider so that it can think more quickly and precisely. The level of education will be affect understanding and assessment of the importance of the environment better the old formal education achieved by respondents in units of years (measured in year).
- c. House Ownership (OWN) is the ownership of place of residence by respondent. Ownership of a house by the respondent will affect the decision respondent affect the willingness of consumers to pay for water quality improvement of PDAM Tirtamarta. Homes categorized as self-owned are stated in a dummy of quality 1, while the house that is still rented is stated in dummy quality 0.

- d. The family members (FAM) of this variable are expected to have a positive effect on the willingness to pay. As more and more the family members, the responsibility to provide better facilities will be higher, therefore the opportunity to be willing to pay will be higher (measured in terms of people).
- e. Water Quality (WQ) is the perception of the quality of existing water definitely will make consumers willing to pay for water quality improvements of PDAM Tirtamarta. Water quality referred to in this research are about the smoothness, cleanliness, and clarity. Water quality that is categorized as good is expressed in dummy of quality 1, whereas categorized qualities that as not good are expressed in dummy quality 0.
- f. Customer Satisfaction (SCUST) customer satisfaction will be one of the decisions for water quality improvements of PDAM Tirtamarta. Customer Satisfaction in question is about the services provided by the PDAM and the credibility of household customers for PDAM. Customer satisfaction is categorized well (in a range 6-10) expressed in dummy 1, whereas consumer satisfaction that is categorized as not good (in a range 1-5) is stated in dummy 0.

## **F. Analysis Method and Data Processing**

This research uses binary logistics regression analysis methods with software tool SPSS which is used to find the relation between variables in research.

## **1. Analysis using Contingent Valuation Method (CVM)**

Contingent valuation method is a direct calculation method, in which case it directly asks willingness to pay (WTP) to the public with emphasis on individual preference assessing public objects that emphasize on standard value of money (Hanley and Spash, 1993).

According to Hanley and Spash (1993), the stages in the CVM can be divided into six parts of work that are as follows:

- a. Build a hypothetical market
- b. Bring up the bid that can be done with open-ended question, bidding game, closed-ended referendum, payment card, dichotomous question.

In this research to know the value that is used dichotomous question. Dichotomous question is to ask respondents, WTP wants to pay for a product by providing a clear bid value. With this method, respondents are asked to vote or not agree on the offer given to the respondent, if the respondent answered "yes" then the bid will be raised to a higher value. Whereas if the respondent replied "no", then bid lower than the first promo.. In this research were asked some respondents with a clear offer on quality improvement that will be used in research.

- c. Suspect the average value of WTP
- d. Suspect the bid curve
- e. Aggregation the data

f. Evaluation

## **2. Binary Logistics Regression Methods**

Willingness to pay of society for quality improvement in this research used binary logistic regression methods. Methods of data analysis used to determine the factors that influence the perception of the individual is a binary logistic regression model.

Logistic Regression is one of statistical models that can be used to analyze patterns of relationship among a set of independent variables and dependent variable using a type of qualitative or categorical. Category of dependent variable can consist of two possible values (dichotomous), such as yes/no, success/failure and others, or more than two values (polychotomous), such as strongly disagree, disagree, agree, and strongly agree (Rosadi,2011).

Binary Logistic regression model used in this research is the regression model with two options that logistic regression with two categories or binomial in the dependent variable “1” if it agreed to pay for quality improvement of PDAM Tirtamarta and “0” if it not agreed to pay for quality improvement of PDAM Tirtamarta.

Binary Logistic regression model forms have two possibilities that consists of a dichotomous variable that concluded of two numbered categories and, in this research using two possibilities which “yes” or “no”

Formulation of the model in full can be outlined as follows:

$$\mathbf{WTP} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{EDU} + \mathbf{b}_2\mathbf{INC} + \mathbf{b}_3\mathbf{OWN} + \mathbf{b}_4\mathbf{FAM} + \mathbf{b}_5\mathbf{WQ} + \mathbf{b}_6\mathbf{SCUST} + \mathbf{b}_7\mathbf{EXP} + e$$

Descriptions:

WTP = Willingness to pay

EXP = Exponent function

$\beta_0$  = Constanta

$\beta_1, \dots, \beta_6$  = Regression coefficient

INC = Family income

EDU = Education Level

OWN = Home ownership

FAM = Number of the family members

WQ = Water quality

SCUST = Consumer Satisfaction

$e$  = error

Statistical used in this research are:

**1. Fitness Model Test**, this test used to assess the provision of predictions using the classification table for calculating the correct and incorrect estimation.

**2. Feasibility Model Test**

a. Negelkerke R Square,

Negelkerke R Square is a test that conducted to find out how big the independent variable able to explain and influence

the dependent variables. The value of the Nagelkerke R Square varies between 1(one) to 0(zero). If the value getting closer to 1 then the model is regarded as goodness of fit, while if the value getting closer with 0 then the model is not considered as goodness as fit. (Ghozali:2011)

In logistic regression, there are no  $R^2$  value as in Ordinary Least Square (OLS) regression. However, Pseudo R-Square can estimate based lack of fit which is indicated by the -2Log-Likelihood. In this case use Nagelkerke Pseudo- $R^2$

$$\text{Nagelkerke Pseudo-}R^2 = \frac{1 - \left[ \frac{-2LL_{null}}{-2LL_k} \right]^{\frac{2}{n}}}{1 - (-2LL_{null})^{\frac{2}{n}}}$$

b. Hosmer-Lemeshow Test (H-L Test)

The feasibility of using H-L Test becomes a null hypothesis that empirical data is appropriate or appropriate to the model (there is no difference between the data so that the model becomes fit)

According to Ghozali (2011), result of H-L Test as follows:

- 1) If equal to or less than 0,05 then rejected hypothesis 0, meaning there is a significant difference between the model with observation value so goodness of fit is not good because of the unpredictability of his observation value.
- 2) If the value of the statistic H-L Test more than 0,05 the hypothesis zero could not be rejected and means the model was able to predict his observation value or can be said to

be acceptable because the model fit with the observation data.

This test aims to research the suitability of the logistic regression model. The Formula of H-L Test are :

$$C^{\wedge} = \sum_{k=1}^g \frac{(O_k - E_k)^2}{V_k}$$

Which are:

$C^{\wedge}$  = Hosmer-Lemeshow Test (H-L Test)

$O_k$  = Observation value at group to-k

$E_k$  = Expectation value at group to-k

$V_k$  = Factor of variant correction at group to-k

### 3. Signification Model Test

#### a. Signification Partial Test

This test using Wald Test (W) that aims to know the existence of  $\beta$  in partial with the following hypothesis.

$H_0 = \beta_i = 0$  (independent variable in partial has no real influence on the dependent variable)

$H_1 = \beta_i > 0$  (independent variable in partial influence to the dependent variable)

For 1,2,3.....n

Statistic test will be used are :

$$W = \left( \frac{\beta_i}{SE(\beta_i)} \right)^2$$

Descriptions :

W = Wald value

$\beta_i$  = Coefficient vectors associated with coefficient of X

SE ( $\beta_i$ ) = Error of  $\beta_i$

So,  $H_0$  is rejected if  $W < 0,05$  which means independent variables influence dependent variable in partial.

b. Simultaneous Model Test (Overall Test)

This test is used to influence the dependent variable against the independent variable at the same time in inside logistic regression model. This test using Likelihood Ratio Test with a hypothesis:

$H_0$  :  $\beta_1 = \beta_2, \dots, \beta_i = 0$  (all dependent variable simultaneously do not influence the dependent variable)

$H_1$  :  $\beta_1 \neq 0$  (there is an independent variable effect on the dependent variable simultaneously)

The statistical test used in this research are as follows:

$$G^2 = -2 \ln \frac{l_0}{l_i}$$

Which are :

$l_0$  = the maximum value of the likelihood of reduced model or explanatory models that consist only of constant (without explanatory variables)

$li$  = the maximum value of the likelihood of full model or model with all independent variable The value of  $G^2$  follow Chi-Squares with degrees free P, so the hypothesis is rejected if  $G^2 < 0,05$  meaning independent variable affects simultaneously