CHAPTER III

RESEARCH METHODOLOGY

A. Research Object

In this study, the author used the data from the Indonesia Family Live Survey (IFLS) that have been surveyed by RAND. The survey of RAND in Indonesia was conducted in 24 provinces in the form of cross-section data, covering the provinces of West Sumatra, Jambi, Lampung, all provinces in Java, Bali, West Nusa Tenggara (NTB), all provinces in Kalimantan, South Sulawesi and West Sulawesi. The IFLS data used in this study was IFLS-5 (2014-2015) released in May 2016 (Strauss et al, 2016).

B. Data Type

The data used in this study was the secondary data obtained by a survey institute called RAND. The other secondary data was obtained in some articles, previous studies, internet, and journals.

The used subjects were individuals in Indonesian households. The data collected from RAND covered 24 provinces in Indonesia. The sample was collected from the household data (RAND) with the total observation of 8,829. The data used cross-section data.

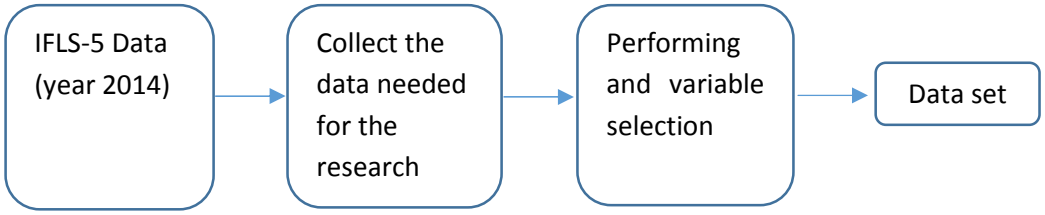
C. Data Collecting Method

1. Sampling Method.

   The research method used the documentation technique from which the study used data from the Indonesia Family Live Survey (IFLS) which was related to the research topic by performing direct cross-section data on IFLS-5 in 2014-2015.

Graphic 3.1

IFLS data collected steps



Resource: Irianti, 2013

D. The Operational Definition of Research Variable

The operational definition is a definition given to a variable or construct by giving meaning or specifying activity or providing an operational needed to measure the construct or variable (Nazir, 1998). In this research the individual happiness or subjective well-being (SWB) is a dependent variable on the model and have four independent variable including income, health, education and unemployment. The operational definition of each variable can be explained as follows:

1. Dependent Variable.

   The dependent variable is a type of variable that explains or is affected by independent variables. In this study, SWB was the dependent variable. The data was taken from individuals in the households (IFLS-5). The research subjects were individuals in households aged 15 years old or older. The SWB variable was obtained from the questionnaires on IFLS-5. The questionnaires that were used in IFLS-5 asked the following question “In the present, do you feel very happy, happy, unhappy, or very unhappy?” The *dummy* variables are described in Table 3.3.

Table 3.1
Dummy

	Positive question (+)		**Negative question (-)**
Alternative answer	Score	Alternative answer	Score
Happy	1	Happy	0
Unhappy	0	Unhappy	1

2. Independent Variable.
   a. Income is the value of individual income in the last one year in form of Rupiah (Rp)
   b. Subjective health status is the individual perception about health in the last one year. The dummy variable is specified 1= health, 0= unhealthy
   c. Education indicates the level of an individual's education taken during his/her life.
   d. Subjective unemployment status indicates the job status of a person in the last month.
   e. Demographic is the category of an individual's living area in the household IFLS-5. The dummy variable is specified 1= city and 0= village.

Table 3.2

Research variable description

Variable	**Variable description**	**Value**
Education	The long of education	Year

continues table 3.2

Research variable description

SHS	The perception about health in last year	1 if a person feels healthy, 0 if a person is unhealthy
SUS	Looking for a job in the last month	Yes or no
Income	The income of member of household	Rupiah (Rp)

E. Quality Data Test

The use of a regression model analysis requires an assumption test to investigate the effect of one variable on other variables. The required assumptions included the normality test, heteroscedasticity, and no multicollinearity.

1. Normality Test

   Normality is a test performed to see the distribution data on the variable used in research. The normality test has a purpose to see any sample data who does not fulfill the normal distribution standard. The normality data can be seen by using *Shapiro Wilk*. In determining the normality test result can be seen by the significant, if the P-value > 0.05 then residual data on the study normally distributed. If the P-value < 0.05, the data is not distributed. Gujarati (2009) said if the normality test is dominant, not normal then the assumption that can be used is the Central

Limit Theorem assumption. The central limit theorem is a condition where the amount of observation is enough (n>30), then the normality assumption can be ignored.

2. Heterocedasticity Test.

The heteroscedasticity test is a test performed to see whether there is an increase in the variance in a dependent variable or not. If the variance of one observation and other observations do not change, it can be said homoscedasticity and if it is different, then it is called heteroscedasticity. If p-value < 0.05, the residual data has heteroscedasticity and if p-value > 0.05, the residual data does not have heteroscedasticity. One of the ways to test the presence of heteroscedasticity in a linear regression model is by performing the Breusch-Pagan test (Cook and Weisberg, 1983).

3. Multicolinearity Test.

The Multicolinearity is used to test the presence of correlation between variable independent and regression model. Ghozali (2007) stated that a good regression model should have a correlation between independent variables. According to Gujarati (2013), when the coefficient of correlation result between dependent variables showed more than 0.8 then it can be said the regression model has a problem with multicollinearity and if the coefficient of correlation result less than 0.8 it can be said the free model from Multicolinearity.

Gujarati (2007) provided some indicators that can be used to see the presence of Multicollinearity on a regression equation:

i. The value of R^2 produces a high estimation model, but many independent variables are not significant and do not affect the dependent variable.

ii. Analyzing a fairly high matrix correlation (generally above 9,0) then it is an indication of multicolinearity.

iii. Sees the value of *tolerance* and the value of *the variance inflation factor* (VIF). A regression is free from the problem of multicollinearity if the value of tolerance is less than 0.1 or equal with the value of VIF ≤ 10

4. Logistic Regression.

The model analysis used in this study was the logistic regression. Logistic regression is one of the multivariate analyses used to predict the dependent variable based on the independent variable.

In the logistic regression, the dependent variable is a variable or categorical dichotomy. The category here can be divided into three conditions. First, the dependent variable has 2 categories, which use binary logistic. Second, when the category is more than 2, then it uses a multinomial logistic regression. Third, the category of the dependent variable is in form of ranking or level, which used the ordinary logistic regression. The prediction in the logistic regression used dichotomy scale, for example: yes and no, good and bad, high and low, etc. Logistic regression establishes an equation or function with the maximum likelihood approach, which maximizes the opportunity to classify the observed object into an appropriate category and convert it into a simple

regression coefficient. Two values that are usually used as the predicted variable include 0 and 1 (for example, 1= high, 0= low). In this study, the dependent variable indicated the possible responses of "1=happy" and "0=unhappy".

In this study, the logistic regression was used to assess the SWB. The estimated model can be seen as follows:

Happiness= f (income, healthy, education)

The model in this research is:

$$ln\left(\frac{\widehat{Subjective\ Well-Being}}{1-\widehat{Subjective\ Well-Being}}\right) = B_0 + B_1 healthy + B_2 logwage + B_3 Education + \varepsilon$$

5. Robust Regression Theory.

Robust regression is known in 1972 and introduced by Andrews. Robust regression is a regression method used when the distribution of errors shows an abnormal or there are some outliers that can affect the model (Olive, 2005). The data has the distribution that shows abnormality and contains an outlier can affect the result of the least squares estimation. Robust regression is an analysis tool used if the data contain an outlier and show the resistant result (Turkan *et al*, 2012). Efficiency and breakdown points have a purpose to explain the level of aggregation of robust distribution. Efficiency has a purpose to show how well a robust estimation is proportional to the least square method without outlier. When the efficiency and breakdown points are higher than the result of an

estimator, the model will be more resistant in analyzing the data containing an outlier.

The method of robust regression is a method to estimate the regression that is insensitive to deviations of underlying assumptions (Gujarati, 2005). M-estimator is a robust regression method that is often used. M-estimator is to estimate the parameter caused by downsizing (outlier).

According to Chen (2002), there are some problems that can be handled using the robust regression technique, including:

i. Outlier problem containing in the dependent variable.
ii. Outlier problem containing in independent variable.
iii. A problem containing in the dependent and independent variable.