CHAPTER III

RESEARCH METHODOLOGY

A. Research Object

In this study, 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 were IFLS-5 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 1175. 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.

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 wage in manufacturing setor is a dependent variable on the model and have five independent variable, including education, experience, age, health status and area of living. 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 wage of manufacturing sector was the dependent variable. The symbol of the dependent variable is Y. Wage is the receipt of labor in the form of money within one month originating from respondents in the manufacturing sector. Wages obtained from the questionnaires on IFLS-5 Book IIIA TK25A1. The manufacturing sector is obtained from the questionnaires IFLS-5 Book 3A

2. Independent Variable.

   Independent variables are those that influence or the cause of changes in the dependent variable (Sugiyono, 2013: 39). In this study the independent variable is the thing that influences the wages of the manufacturing sector labor. The independent variables in this study are education level (X1), experience (X2), age (X3), health status (X4) and area of living (X5). The explanation of each independent variable as follows:

a. Level of education (X1)

Education level is a variable implies the level of knowledge possessed by labor based on the level of education completed. The indicator used to measure education levelis the level of formal education taken by the workforce from not passing Elementary School to higher education. Level education is obtained from the questionnaires IFLS-5 Book K AR16 and AR17. The following table will be presented regarding levels education used in this study.

Table 3. 1
Level of Education

No	The Highest Level of Schooling Attended
1.	No school
2.	Elementary school
3.	Junior high school
4.	Senior high school
5.	Diploma
6.	University

b. Experience (X2)

Work experience is obtained from the age of the workforceminus the length of years of successful education takenminus the age of first time entering school. Work experience is expressed in units of years.

Experience is obtained from the questionnaires of IFLS-5 Book TK 23A2.

c. Age (X3)

The age of a person can be known if the date, month and the year of birth is known. The age of a person is measured according to the last birhday, age of respondents in this study minimum of 15 years and above. The age variable is obtained from IFLS-5 Book T2 AR 09.

d. Health Status (X4)

Subjective health status is the individual perception about health in the last one year. The dummy variable is specified 1= health, 0= unhealthy. Health status is obtained from the questionnaires of IFLS-5 Book IIIB KK01.

e. Area of Living (X5)

The area of living or domicle is the place that is used as the center of his residence. This variable is expressed in the form of a dummy, namely urban code 1 and rural code 0. Domicile is obtained from the questionnaires of IFLS- 5 Book T SC 05.

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 some 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 are 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. Robust Regression Theory

Robust Regression Theory In 1972 Andrews introduced robust regression. Regression is a regression method that is used if the distribution of errors shows abnormal or there are several outliers that affect the model (Olive, 2005). If the data have a distribution that shows abnormalities and contains an outlier so that the outlier can affect the result of the least squares estimation. Robust regression is a tool for analyzing data containing outline and showing results that are resistant where there are outliers (Turkan et al, 2012). Efficiency and breakdown points aim to explain the level of robustness of the robust estimation results. An efficiency that aims to show how well the results of a robust estimate are proportional to the least squares method without outliers. If the higher efficiency and breakdown points of the results of an estimator, the more resistant a model is in analyzing data that has outliers.

Robust regression method according to (Gujarati, 2005), which is a method for estimating regression coefficients that are not sensitive to deviations from the underlying assumptions. M-estimator is a robust regression method that is often used. The M-estimator is considered to be good for estimating parameters caused by outliers. Some problems that can be handled using robust regression techniques, according to Chen (2002), namely:

1. Outlier problems found in the dependent variable.
2. Outlier problems found in the independent variable.
3. Outlier problems found in both, namely the dependent variable and the independent variable.

F. Data Analysis

A steps in this study used Stata 14 for helping data analysis as follows :

1. Multiple Linear Regression

This study used multiple linear regression technique, which connecting between variables by including the element of causality. The model in the regression adopts The Mincerian model. The research variables equations as follows :

$\beta lnW_i=\beta_0+\beta_1 Edu_i+\beta_2 Exp_i+\beta_3 Age_i+\beta_4 Healthstatus_i+\beta_5 Areaofliving+\varepsilon_i$

Known :

W_i = individual wage $_i$

Edu_i = individual school year $_i$

Exp_i = individual work experience $_i$

Age_i = individual of age

Healthstatus = indiviual health status (unhealthy = 0, healthy = 1)

Areaofliving = individual area of living (rural = 0, urban = 1)

ε_i = error term

In this case of study school year replaced to a level of education other than that independent variables namely age, health status, and area of living. So the equations model as follows:

$$lnY_{it} = \beta 0 + \beta_1 Es_{it} + \beta_2 JHS_{it} + \beta_3 SHS_{it} + \beta_4 Diploma_{it} + \beta_5 Univ_{it} + \beta_6 Exp_{it} + \beta_7 Age_{it} + \beta_8 Health\ Status_{it} + \beta_9 Area\ of\ Living_{it} + \varepsilon_i$$

LnY = log a wage in manufacturing sector

Es = a level of education Elementary School (ES=1, other = 0)

JHS = a level of education Junior High School (JHS = 1, other =0)

SHS = a level of education Senior High School (SHS = 1, other = 0)

Diploma = a level of education Diploma (Diploma = 1, other = 0)

Univ = a level of education University (University = 1, other = 0)

Exp = work experience in proxy with the age minus a year of school minus year of t

Age = Age in one year

Health Status = Health status (healthy = 1, unhealty = 0)

Area of Living = area of living (urban = 1, rural =0)

ε_i = error term

B = regression coefficient