CHAPTER THREE
RESEARCH METHODOLOGY

A. Object / Subject of Research

This study is aimed at districts / cities in Riau Province. Pekanbaru municipal elections as the study area due to facilitate the collection of data, in addition Pekanbaru is the capital of Riau province.

B. Types and Sources of Data

The data used in this study are primary and secondary quantitative method which is the time series data from 2010-2014. About the Local Revenue (PAD), Total Population, Government Spending and Gross Regional Domestic Product (GRDP) is obtained from the Office of the Department of Revenue, the Central Statistics Agency (BPS) Riau, literature-literature / books and reports related to this writing.

C. Data Collecting Technique

Data collecting was done in this study by way author proposes permit research to agencies associated with this author in a study area, after being given permission research and obtain the data required, and then the data will be processed and used as materials analysis to prove the hypothesis that has been put forward.

D. The Variables Operational Definition of Research

In this study consisted of the dependent variable [Y] is Local Revenue (PAD) of Riau Province, Independent Variables [X_1] is Total Population, [X_2] Government Spending, and [X_3] Gross Regional Domestic Product (GRDP) of Riau Province.

Local Revenue (PAD) is a source of local revenue derived from the results of local taxes, levies result, the results of companies belonging to the region and other areas of wealth management results were separated and others Local Revenue (PAD) legitimate in Riau Province. Unit in thousand rupiah, data is taken from the Central Bureau of Statistics Riau Province 2010-2014.

Total Population is the number of people living in the province of Riau. Data is taken from the Central Statistics Agency of Riau Province 2010-2014.

Government Spending is the expenditure including the government sector and purchases of goods and services used for the payment of subsidies which implements the functions of government. Government expenditure data is taken from the Central Statistics Agency of Riau Province 2010-2014.

Gross Regional Domestic Product (GRDP) is the total value of production of goods and services produced in the region (regional) within a certain specified time (one year) in the province of Riau. Data is taken from the Central Statistics Agency of Riau Province 2010-2014.

E. Analysis Method

Data analysis technique used quantitative approach which quantitatively explains the object of study is based on facts and data. Then, analyze descriptively to answer the problem of research. This approach is expected to explain the condition of the object under study to look at the Total Population, Government Spending in the form of direct and indirect spending and Gross Regional Domestic Product at constant prices of the Local Revenue in districts/cities in Riau Province.

This study uses panel data. Panel data regression is used to answer the purpose of this study is to determine whether the number of Total Populations, Government Spending and Gross Regional Domestic Product affect Local Revenue in districts/cities in Riau Province. Panel data is a combination of time series data (time series) and data cross (cross section). There are several advantages of using panel data in economic research. According to Gujarati (2010) in (Tobing, 2015), the advantage of using panel data is as follows:

1. Data Panel is able to accommodate the level of heterogeneity of the variables that are not included in the model.

2. Data Panel able to indicated and measured the effect is the same and cannot be obtained with a simple cross section pure or pure time series.

3. Data Panel was able to reduce co linearity between variables.

4. Data Panel can test and build more complex behavioral models.

5. Data Panel can minimize bias generated by individual regression because more data units.

The data used in this panel data is a combination of cross section data and time series. Time series data used are annual data for 5 years from 2010 to 2014, and the cross section consists of 12 districts/cities in Riau Province which consists of Kuantan Singingi, Indragiri Hulu, Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu, Rokan Hilir, Meranti Islands, Pekanbaru, Dumai.

Panel Regression models are as follows:

$$Y = \alpha + b_1 X_1 + b_2 X_2 + b_3 X_3 + e$$

Where:
Y=   Dependent Variable (LDR)
α =   Constants
X1= 1st Independent Variable
X2= 2nd Independent Variable
X3= 3rd Independent Variable
b{1,2,3} = Regression Coefficient of each Independent Variable
e =   Error term

In the regression model estimation method using panel data can be performed through three approaches:

a. Common Effect Models

Panel data model approach is the simplest because only combines the data time series and cross section. In this model neglected dimension of time as well as individuals, so it is assumed that the behavior of corporate data together in different periods. This method can use the approach Ordinary Least Square (OLS) or a least squares technique for estimating panel data model.

The regression equations in the model Common Effect can be written as follows:

$$Y_{it} = \alpha + X_{it}\,\beta + \varepsilon_{it}$$

Where:

i = Kuantan Singingi, Indragiri Hulu,…. Dumai

t = 2010, 2011, 2012, 2013, 20214

Where i indicate the cross section (people) and t is the time period. Assuming the error component in the processing of ordinary least squares, estimation process separately for each unit cross section can be done.

b. Fixed Effect Models

This model estimates that the differences between individuals can be accommodated on the difference intercept. To estimate the Fixed Effects panel data models using the technique of dummy variables to capture the difference between the company's intercept, the intercept differences can occur due to differences in the work culture, managerial, and incentives. Nevertheless, it slopes equally between the companies. The estimation model is often called the Least Squares Dummy Variable technique (LSDV).

In the Fixed Effects models, each individual is an unknown parameter and estimated by using dummy variables as follows:

$$Y_{it} = \alpha + i\alpha + X'_{it}\,\beta + \varepsilon_{it}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_n \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha \\ \alpha \end{bmatrix} + \begin{bmatrix} i & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & i \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_n \end{bmatrix} + \begin{bmatrix} x_{11} & x_{21} & x_{p1} \\ x_{12} & x_{22} & x_{p2} \\ x_{1n} & x_{2n} & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_n \end{bmatrix}$$

These techniques are called Least Square Dummy Variable (LSDV). In addition to in addition be applied to each individual effect, LSDV can also

accommodate a systemic effect. This can be done through the addition of a dummy variable in the model.

c. Random Effect Model

This model will estimate the panel data where possible disturbance variables are interconnected across time and between individuals. In the Random Effects Model intercept differences are accommodated by the error terms of each company. The advantages of using a model are remove heteroscedasticity of Random Effect. This model is also called the Random Error Component Model (ECM) or technique Generalized Least Square (GLS).

Random Effects Model equation can be written as follows:

$$Y_{it} = \alpha + X'_{it}\beta + w_{it}$$

> i = Kuantan Singingi, Indragiri Hulu,…. Dumai
> t = 2010, 2011, 2012, 2013, 2014

Where:

$w_{it} = \varepsilon_{it} + u_1$; $E(w_{it}) = 0$; $E(w_{it}^2) = \alpha^2 + \alpha_u^2$;

$E(w_{it'}w_{jt-1}) = 0; 1 \neq j$; $E(u_{i'}\varepsilon_{it}) = 0$;

$E(\varepsilon_{i'}\varepsilon_{is}) = E(\varepsilon_{it'}\varepsilon_{jt}) = E(\varepsilon_{it'}\varepsilon_{js}) = 0$.

Although, the error components of *wt* are homoscedastic, in fact, there is a correlation between the *wt* and *wit-s* (equicorrelation), namely:

$$\mathrm{Corr}(w_{it'}\, w_{i(t-1)}) = \alpha_u^2 / (\alpha^2 + \alpha_u^2)$$

Therefore, the OLS method cannot be used to obtain an efficient estimator for random effects models. An appropriate method for estimating the random

effects models is the Generalized Least Squares (GLS) assuming homocedastic and no cross sectional correlation.

F. Selection of Model

For the selection of the right model to manage the data panel, can be tested as follows:

1. Chow Test

   Chow test is a test to determine the model Fixed Effect or Random Effect most appropriately used in estimating panel data.

2. Hausman Test

   Hausman test is a statistical test to select whether the model Fixed Effect or Random Effect most appropriately used.

3. Lagrange Multiplier Test

   To determine whether Random Effect Model is better than Common Effect Method (OLS), test was used Lagrange Multiplier (LM).

After obtaining the right model, the regression results of the model is to prove the hypothesis the presence or absence of significant influence then tested the significance of the t test and F test. In the test specification models in the study, the authors used several methods:

a. Chow Test

   Chow Test is a test to determine the model Fixed Effect or Random Effect most appropriately used in estimating panel data. The hypothesis of the Chow Test is:

H0:   Common Effect Model or pooled OLS

H1:   Fixed Effect Model

Basic rejection of the above hypothesis is by comparing the calculation of the F-statistic with F-table. Comparison is used if the results of the F count is greater (>) of F table then H0 rejected, which means the most appropriate model used is the Fixed Effects Model. Vice versa, if F count is smaller (<) of F table then H0 is accepted and the model used is Common Effect Model (Widarjono in Basuki and Yuliadi, 2015).

Calculation of F statistics obtained in Chow Test formula (Baltagi in Basuki and Yuliadi, 2015):

$$F = \frac{\dfrac{(SSE_1 - SSE_2)}{(n - 1)}}{\dfrac{SSE_2}{(nt - n - k)}}$$

Where:
SSE$_1$:   Sum Square Error from Common Effect Model
SSE$_2$:   Sum Square Error from model Fixed Effect
n   :   Number of Companies (Cross Section)
nt   :   Number of cross section x Number of time series
k   :   Number of Independent Variables

While F table obtained from

$$F\text{-tabel} = \{\alpha : df (n\text{-}1, nt\text{-}n\text{-}k)\}$$

Where:

α:	The Significance level used (Alfa)
n:	Number of Companies (cross section)
nt:	Number of cross section x Number of time series
k:	Number of Independent Variables

b. Hausman Test

After completing the Chow test and obtained the right model is Fixed Effect, then the next we will examine which model among models Fixed Effect or Random Effect most a

ppropriate, this test is referred to as Hausman test.

Hausman test can be defined as statistical tests to select whether the model Fixed Effect or Random Effect most appropriately used. Tests conducted by the Hausman test the following hypotheses:

H0:   Random Effect Model

H1:   Fixed Effect Model

Hausman Test will follow the distribution of *Chi-squares* as follows:

$$m = \hat{q} \, Var \, (\hat{q}) - 1 \, \hat{q}$$

Hausman test statistic follows the *Chi Square* statistic distribution with a *degree of freedom* as much as k, where k is the number of independent variables. If the value of the Hausman statistic is greater than the critical value, H0 is rejected and the right model is a model Fixed Effect while conversely if Hausman statistic value is smaller than the critical value, the appropriate model is the model of *Random Effect.*

If the Hausman test showed no significant difference (p> 0.05), it reflects that the random effects estimator is not free safe free of bias, and therefore more advisable to estimate fixed effect rather than an effect estimator remains.

c. Lagrange Multiplier Test

Lagrange Multiplier (LM) is a test to determine if the Random Effect Model or Common Effect Model (OLS) is most appropriately used. Random Effect significance test was developed by Breusch Pagan. Breusch Pagan method for Random Effect significance test is based on the residual value of the OLS method.

The value of LM statistics is calculated based on the following formula:

$$\mathbf{LM} = \frac{nT}{2(T-1)} \left[\frac{\sum\limits_{i=1}^{n} (T\,\hat{e}\,i)^2}{\sum\limits_{i=1}^{n} \sum\limits_{t=1}^{T} \hat{e}_{it}^2} - 1 \right]^2$$

Where:
- n =   Number of Individuals
- T =   Number of Time Periods
- e =   residual method of *Common Effect* (OLS)

Hypothesis is:

    H0: Common Effect Model

    H1: Random Effect Model

LM test is based on the distribution of *Chi-Squares* with *a degree of freedom* for the number of independent variables. If the value of LM statistic is greater than the critical value of chi-squares then we reject the null hypothesis, which means precise estimation for panel data regression model is

a method of Random Effect on Effect Common methods. Conversely, if the value of the LM statistic is less than the value of chi-squares as critical value, then we accept the null hypothesis, which means that the estimates used in panel data regression method is not a method Common Effect Random Effect (Widarjono, 2013).

G. Quality Test Data

By using Ordinary Least Square method (OLS), to produce the value of the model parameter estimators are more precise, it is necessary to detect whether the model deviates from the classic assumption or not, the detection consists of:

1. Multicollinearity Test

Multicollinearity can be defined as a situation where one or more independent variables collinear expressed as a combination of other variables. This test aims to determine whether the regression found a correlation between the independent variables. If there is a correlation, then it called the problem of multicollinearity. One way to detect the multicollinearity is:

a. R^2 is quite high (0.7 to 0.1), but the t-test for each regression coefficient is not significant.

b. High-R^2 is Low sufficient condition (sufficient) but not a necessary condition (Necessary) for the occurrence of multicollinearity, because in R^2 is low (<0.5) can also occur multikolinearity.

c. Regressing the independent variables X with another variables independent, then in his R^2 calculated by F test:

1) If F^* > F table, means H_0 is rejected, so then, there is multicollinearity

2) If F^* < F table, means H_0 is accepted, so then, there is no multikolinearity

One characteristic symptom of multicollinearity is the model has a high coefficient of determination (R^2) say above 0.8 but few significant independent variables affect the dependent variable through the t test. However, based on the F test are statistically significant, which means all the independent variables jointly affect dependent variable (Widarjono, 2013).

To resolve the problem, multicollinearity, the independent variables were correlated with other independent variables should be removed. In the case of GLS method, this model has been anticipated from multikolinearity.

2. Heteroscedasticity Test

A regression model is said to be exposed in the event of inequality heteroscedasticity residual variance from one observation to another observation. If the variance of the residuals and the observations of the other observations remained, then it called homoscedasticity. If the variance is different, it is called heteroscedasticity.

Heteroscedasticity can occur if there is some data loner (outliers) in the model.

One of the test statistics that can be used to test whether the variance of the error is homoscedasticity or not is to test the Park to assume that σ_i^2 is a

function of the explanatory variables X. The equation used to indicate the relationship is:

$$\sigma_i^2 = \sigma \, X_i^\beta e^{Vi} \quad \text{or}$$

$$\ln \sigma_i^2 = \ln \sigma + \beta X_i + V^i$$

Because σ_i^2 is unknown, so it is estimated with e_i^2 , then the model becomes

$$\ln e_i^2 = \alpha + \beta X_i + V^i$$

If β significantly, it means there is heteroscedasticity.

So, there are two stages in the Park test, namely:

a. Regressing Y to X by the least squares method, as well as get e_i with e_i^2.

b. Regressing e_i^2 to X using the model $\ln e_i^2 = \alpha + \beta X_i + V^i$ (Setiawan dan Kusrini, 2010).

This test aims to test whether the regression model was occurred inequality variance of residuals from one observation to another observation. If the variance of the residuals of an observation to observation of others still, it is called heteroscedasticity. A good regression model is the absence of heteroscedasticity. In the case of GLS method, this model has been anticipated from heterocedasticity. Detection of heteroscedasticity:

1) If there is a particular pattern, such as dots form a pattern that is specific manageable units (corrugated, widened and then narrowed), there have been heteroscedasticity.

2) If there is no clear pattern, as well as the points spread above and below the number 0 on the Y axis, it does not happen Heteroscedasticity.

H. Statistical Analysis Regression Test

Significance test is a procedure used to examine errors or correctness of the results of the null hypothesis of the sample.

1. Coefficient of Determination Test (R-Square)

In essence, the coefficient of determination R^2 measures how far the ability of the model is to explain variations in the independent variables to measure the goodness of a model (goodness of fit). That is how the regression of established value according to the data. If all data is placed on the regression line or in other words all the residual value is zero then we have the perfect regression line. And this is very rarely the case. Generally what happens is $\hat{e}_i$ can be positive or negative. If this is the case, means regression line is not one hundred percent perfect. However, the hope is to get the regression line that led to EI as small as possible. The determination coefficient values between 0 and 1 ($0 < R^2 < 1$), the value (R^2) is small means the ability of the independent variables in explaining the variation of variables is very limited. A value close to 1 means that the independent variable provide almost all the information needed to predict the variation dependent models (Gujarati in Awanis,2015).

The use of R Square (R squared) often cause problems, namely that its value will always increase with the addition of independent variables in a model. This will lead to bias, because if you want to obtain a model with high

R, a researcher can arbitrarily adding independent variable and the value of R will increase, regardless of whether the additional independent variables associated with the dependent variable or not. Because of the weaknesses in the calculation of R^2, many researchers are advised to use the Adjusted R Square. Interpretation is the same as R Square, but the value of Adjusted R Square can rise or fall with the addition of a new variable, depending on the correlation between the additional independent variables with the dependent variable. Adjusted R Square value may be negative, so if the value is negative, then the value is considered to be 0, or not at all independent variables able to explain the variance of the dependent variable (Muttaqin dkk,2014).

2. F-Statistics Test

F-statistic test is done to see how much influence the independent variable as a whole or together on the dependent variable. The steps are performed in this test as follows:

a. Formulating the Hypothesis

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta4 = 0$, meaning that together there is no influence of independent variables on the dependent variable.

H_a: $\beta_1 = \beta_2 = \beta3 = \beta_4 \neq 0$, meaning that together the influence of independent variables on the dependent variable.

b. Decision-Making

Decision-making in the F test is done by comparing the probability of simultaneous effect of independent variables between the independent

variable on the dependent variable with a value of alpha used in this study the authors used an alpha of 0.05.

If the probability of independent variables> 0.05, then H0 accepted, it means the independent variables simultaneously (together) do not significantly affect the dependent variable.

If the probability of independent variables <0.05, then H0 is rejected or accept Ha, meaning independent variables simultaneously (together) influence on the dependent variable.

3. T-Statistics Test (Partial Test)

The t-test was conducted to see the significance of individual independent variables on the dependent variable to consider other independent variables are constant. The steps are performed in this test as follows:

a. Formulating the Hypothesis

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta4 = 0$, it means, no individual influence of independent variables on the dependent variable.

H_a: $\beta_1 = \beta_2 = \beta3 = \beta_4 \neq 0$, it means, that no individual influence of independent variables on the dependent variable.

b. Decision-Making

Decision-making in the T test is done by comparing the probability of the independent variable on the dependent variable with a value of alpha used in this study the authors used an alpha of 0.05 (5%).

If the probability of independent variables> 0.05, then H0 accepted, it means the independent variables too partially (own) did not significantly affect the dependent variable.

If the probability of independent variables <0.05, then H0 is rejected or accept Ha, meaning partially independent variables (itself) influence on the dependent variable.

This test can be done by comparing t arithmetic with t table. The formula to get the t is as follows:

$$t = (bi - b)/sbi$$

Where:
  Bi =   coefficients of independent variables to-i
  b =   the value of the null hypothesis
  sbi =   standard deviation of the independent variables to-i

At the significance level of 5 percent with testing criteria are used as follows:

1) If t <t table then H0 is accepted and H1 rejected, which means that one of the independent variables (independent) does not affect the dependent variable (dependent) Significantly.

2) If t> t table then H0 rejected and H1 accepted, which means that one of the independent variables (independent) affects the dependent variable (dependent) significantly.