CHAPTER III

RESEARCH METHODOLOGY

3.0. Basic Concepts and Theories

This section includes the analytical and methodological framework of the study.

3.0.1. Multidimensional Scaling Analysis

Multidimensional scaling (MDS) analysis is a classical multivariate and exploratory data analysis under the ordination technique used in information visualization that attempts to arrange objects in a space with a particular number of dimensions so as to reproduce the observed measurement of the similarity or dissimilarity of a pair of objects called proximities. The goal of the analysis is to detect a meaningful underlying dimension that explains the observed proximities between the investigated objects or to display the information contained in a distance matrix. It is a process used to reveal the structure of a dataset by plotting points in two or more dimensions. Generally, it is a family of procedures for constructing a spatial model of objects using information about the proximities between the objects. MDS algorithm aims to place each object in N-dimensional space such that the between-object distances are preserved as well as possible. There were three (3) stages of MDS analysis:

1. First, a distance matrix is generated for all objects. A distance matrix is a squared matrix containing the distances, taken pairwise, between the objects of a set.
2. Second, proximities are extracted from the Euclidean distance matrix based on the distance coefficients to obtain the coordinate of the object.
3. Lastly, coordinates are plotted into N-dimensional graph of scaled distances.

3.0.2. Multidimensional Scaling (MDS) Method

A classical MDS method is a ***KxK*** symmetric matrix **(Δ)** of proximity among k objects represented by the i^{th} row and j^{th} column shown by the cell entry d_{ij}. $\boldsymbol{d_{ij} = D = d(p,q) = d(q,p)}$ is the Euclidean distance calculated between the points representing objects i and j in n-dimensional space, and from the entries in the i^{th} and j^{th}:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

Create a double-centered matrix of **(Δ)**, designated as **(Δ*)**, using Torgerson's model:

$$\Delta^*_{ij} = -0.5(\delta^2_{ij} - \delta^2_{i.}\ \delta^2_{.j} - \delta^2_{..})$$

Perform an eigendecomposition on **(Δ*)**:

$$(\Delta^*) = \mathbf{VA^2V'}$$

where:

V = **K**x**Q** matrix of eigenvectors,

$\mathbf{A^2}$ = **Q**x**Q** diagonal matrix of eigenvalues,

Q is the rank of **(Δ*)** usually equal to k

to obtain the matrix of coordinates, ***X***, from the first m eigenvectors ($\mathbf{V}_m$) and first m eigenvalues ($\mathbf{A^2}_m$):

$$\mathbf{X} = \mathbf{V}_m\, \mathbf{A^2}_m$$

such that the interpoint distances has a least-squares fit to the entries in **(Δ)**.

For a graphical assessment, scree plot is used as a plot of the reproduced distances for a particular number of dimensions against the observed input data. This plot shows the reproduced distances plotted on the vertical *Y*-axis versus the original proximities plotted on the horizontal *X*-axis. This plot also shows a step-function. If all reproduced distances fall onto the step-line, then the rank-ordering of proximities would be perfectly reproduced by the respective MDS solution. Deviations from the step-line indicate lack of fit. The scree plot shows the eigenvalues plotted in the order that they were factored from the proximities matrix. Eigenvalues are related to the variance in the double-centered distances that is explained by each eigenvector.

The stress test is used to evaluate how well or poorly a particular configuration reproduces the observed distance matrix. It is the sum of first *m* eigenvalues relative to sum of all *q* eigenvalues (usually $q = k$):

$$\text{Stress} = \sqrt{\frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})^2}{\sum_{ij} d_{ij}^2}}$$

Eigenvalues measure variance associated with each dimension of the MDS solution. The smaller the stress value, the better is the fit:

Stress-value	Goodness-of-fit
0.200	poor
0.100	fair
0.050	good
0.025	excellent
0.000	perfect

3.1. Methodology

The researcher used the following method:

3.1.1. Data Source

Data from the official website of Philippine Statistics Authority (PSA) were retrieved in the Regional Quickstat Editions report. The independent variables that were considered in the study are the 10 major economic indicators for monitoring economic progress. Data on the major economic indicators from the year 2006 to 2015 were considered. The dependent variables are the 17 regions in the Philippines.

A. The major economic indicators under analysis are:

a. Birth Rate (**BIRTH**). The birth rate is the total number of live births per 1,000 of a population in a year or period.

b. Death Rate (**DEATH**). The death rate is the ratio of deaths to the population of a particular area during a particular period of time, usually calculated as the number of deaths per one thousand people per year.

c. Employment Rate (**EMP**). The employment rate is the percentage of the labor force that is employed.

d. Gross Domestic Product (**GDP**). The gross domestic product (GDP) is the monetary value of all the finished goods and services produced within a region's or country's borders in a specific time period.

e. Inflation Rate (**INFLA**). The inflation rate is the rate at which the general level of prices for goods and services is rising and the purchasing power of currency is falling.

f. Labor Force Participation Rate (**LABOR**). The labor force participation rate is the percentage of people employed and self-employed plus unemployed but ready and able to work.

g. Purchasing Power of the Peso (**PESO**). The purchasing power of the peso is the value of a peso expressed in terms of the amount of goods or services that one unit of money can buy.

h. Population (**POP**). The population (in million) of the country consists of all persons falling within the scope of the census. The total may comprise either all usual residents of the country or all persons present in the country at the time of the census.

i. Underemployment Rate (**UNDEREMP**). The underemployment is a measure of employment and labor utilization in the economy that looks at how well the labor force is being utilized in terms of skills, experience, and availability to work.

j. Unemployment Rate (**UNEMP**). The unemployment rate is a measure of the prevalence of unemployment.

B. The regions in the Philippines are:

a. **NCR** -National Capital Region

b. **CAR** -Cordillera Administrative Region

c. REGION 1 (**R1**) -Ilocos Region

d. REGION 2 (**R2**) -Cagayan Valley

e. REGION 3 (**R3**) -Central Luzon

f. REGION 4A **(R4A)** -Southern Tagalog -CALABARZON

g. REGION 4B **(R4B)** -Southern Tagalog -MIMAROPA

h. REGION 5 (**R5**) -Bicol Region

i. REGION 6 (**R6**) -Western Visayas

j. REGION 7 (**R7**) -Central Visayas

k. REGION 8 (**R8**) -Eastern Visayas

l. REGION 9 (**R9**) -Zamboanga Peninsula

m. REGION 10 (**R10**) -Northern Mindanao

n. REGION 11 (**R11**) -Davao Region

o. REGION 12 (**R12**) –SOCCSKSARGEN

p. REGION 13 (**R13**) -CARAGA

q. **ARMM** -Bangsamoro

3.1.2. Statistical Index Construction

There were four (4) phases in formulating a statistical index:

A. Data Standardization

The major economic indicators were expressed in various units and have to be normalized to make these indicators comparable. Standardization addresses the measurement units by converting each indicator into a common variable scale. Standardization avoids introducing distortions stemming from differences in variable means (Freudenberg, 2003). The standardization formula is:

$$Z = \frac{X - \mu}{\sigma}$$

where:

x = is the value of indicator p for region c that is being standardized

μ = is the average value of indicator p for region c

σ = is the standard deviation of indicator p for region c

B. Weighing Method

The normalized indicators were subjected to multidimensional scaling (MDS) analysis to reduce the number of dimensional space into 2 dimensions while retaining as much as possible of the variation present in the data set. The identified variables with high coordinates were considered in the final construction of the statistical index.

C. Aggregation Method

Aggregation method was applied to transform the weighted normalized variables into a composite index, the **Economic Development Index** (**EDI**), by collecting all the highest coordinates between the 2 dimensions of every variable and calculate the mean. It was clearly done with a view to improve the overall capacity of the economic indicators as variables. (The Economist Newspaper Ltd., 2006).

EDI = [(VARIABLE 1)+(VARIABLE 2)+(VARIABLE 3)+…+(VARIABLE N)] / [N]

D. Validation Method

Since index values can easily become distorted if one value is much less or much more significant than the others (The Economist Newspaper Ltd., 2006), bootstrap resampling technique was then used to evaluate its statistical properties such as unbiasedness, precision, accuracy, and consistency. There were three (3) stages of Bootstrap resampling:

1. Sample Iteration. Iterate the indices by sample size *s* (25%, 45%, and 75%) using the following code:

```
=INDEX(SAMPLES,ROWS(SAMPLES)*RAND()+1,COLUMNS(SAMPLES)*RAND()+1)
```

2. Bootstrapping. Apply different bootstrap resampling *B* (200, 400, 600, 800, and 1000) in every iterated sample size.

3. Table and Graph. Tabulate and graphically illustrate the bias and the standard error of bootstrap results.

3.1.3. Economic Development Index (EDI) Ranking of Philippine Regions

Ranking of Philippine regions was based from the highest to lowest computed Economic Development Index (EDI) values with the computed mean index as the average rank for the Philippines.

3.1.4. Human Development Index (HDI) Ranking Presentation

Presentation of rankings was made with the ranking of provinces for the Human Development index (HDI).