CHAPTER II

LITERATURE REVIEW

2.1 Model and its Role in Transportation Planning

The model is a simplified representation of an actual situation. The model can be used to investigate the consequences of a policy or a certain strategy. The model has the following functions (Bell, MGH; Bonsall, PW; Leake, GR 1997):

1. Predict future conditions beyond policy interventions.

2. Predict future conditions by assuming that any part of a particular policy or design can be realized.

3. Test the achievements of the policy interventions provided, in each future set imagined.

4. Generate short-term forecasting as part of a related control system as can be found in a sophisticated traffic control system.

2.2 Model in Transportation Planning

Transportation planning usually uses a 4-shelf transportation planning model (four-step sequential demand modeling) that is:

1. Trip generation.

2. Trip distribution.

3. Mode division.

4. Trip assignment.

In this 4-step process, the results of each step become input to the next step,

which also uses input that matches the characteristics of alternative plans in the study (network description) and from the land use phase and socioeconomic factors (Papacostas, CS, Prevedouros, PD 2001).

The trip generation phase predicts the number of movements to be carried out by a person in each zone of origin by using detailed data on the level of trip generation, socioeconomic attributes, and land use (Tamin 2003). The need for movement of both generation and attraction is the result of trip generation which is the first stage of transportation planning.

2.3 Factors Affecting Movement Needs

Passenger demand for transportation is related to regional population and individual motivation to travel or their propensity to travel. In addition socioeconomic activities and factors that support travel and the availability of services and related infrastructure also correlate with passenger demand.

At the local and regional level, socioeconomic or demographic variables along with direction and economic growth will play a major role in determining the number of passengers in an area (Ashford, NJ; Mumayiz, S; Wright, PH; 2011). Referring to the research conducted by (Fretes 2014) socioeconomic factors which are considered to influence the movement of passengers are as follows:

1. Total population.

2. Human development index.

3. Number of lecturers.

4. Education index.

5. The number of domestic tourists.

6. Number of foreign tourists.

7. Number of motorized vehicles.

8. GRDP at current prices.

9. Population purchasing power.

10. Per capita income of the population.

11. Per capita population expenditure.

2.4 Train

Corresponding to *Undang Undang* No.23 of 2007, a train is a railroad vehicle with movable power, both traveling alone and coupled with other railroad facilities, which will or are currently moving on railroad tracks related to railroad travel.

According to Warpani, initially the term train known in Indonesia arose because in the past the fuel used was coal or wood, so that when train was running it emitted clouds of smoke from the chimney in addition it also carried quite a lot of sparks (Ferawati 2014).

Trains can be distinguished according to their respective characteristics, the following are the types of trains that are distinguished from their characteristics, among others (Ferawati 2014):

1. Ordinary train, is a train whose journey is written in the train travel chart, written in the time and running list every day specified in the graph and in the time list.

2. Facultative train, is a train whose trip is not written in the train travel chart and written in the time list but only runs if needed.

3. Extraordinary trains, are trains whose journeys are not written in the train travel chart and are not written in the time list but are determined according to the requirements.

   Train modes have advantages and disadvantages in carrying out their functions as one of the modes of transportation for goods and people. The benefits of railroad transportation can be explained, among others:

1. Railroad transportation mode is a type of transportation mode that provides people / goods services in short, medium and long distances with large capacity (mass transportation).

2. The energy used is relatively small, even using battery-powered electricity from a power source that allows energy-efficient use.

3. Providing quite high time related to relative speed and safety of the trip will be better than other modes, because it has a track (track) and complete terminal facilities.

4. The total variable cost (operational cost) per day calculation is quite high, but the variable cost per ton per km is very low (because the transport capacity is quite large) compared to mode development.

   In its advantages, trains also have disadvantages including:

1. Requires special facilities and infrastructure that cannot be used by other modes of transportation, as a consequence it is necessary to provide special transportation equipment (carriages and locomotives).

2. The investment spent is quite high because the train requires special treatment in the maintenance process.

3. Services of people / goods are only limited to the path (not door to door).

4. If there are obstacles (accidents) on the route, then it cannot be diverted immediately to other lanes.

2.5 Train Station

The station is where the train departs or stops to serve the ups and downs of passengers and loading and unloading of goods. In addition, the station also functions as a train traffic controller. Large stations often also serve as train and locomotive maintenance. During the train journey through many stations but not stopped, these stations have the duty to give signals and regulate the smooth operation (Ferawati 2014).

2.6 Theoretical Basis

Simple modeling is done by identifying the level of relationship between the dependent variable with the independent variable. The number of train passengers at Yogyakarta Tugu Station in this study was chosen as the dependent variable. While the independent variable consists of several socioeconomic factors that are considered to influence the dependent variable that has been determined. The socioeconomic factors are as follows:

2.6.1 Total Population

The population shows the number of people who live in a certain area for 6 months or more and or those who live less than 6 months but intend to stay. Total Population acts as X_1 in the calculation.

2.6.2 Human Development Index

The Human Development Index (HDI) or often referred to as the Human Development Index (HDI) is a measurement scale that is able to show the performance of human development in an area in a certain time. HDI can be used to compare between regions and between times. The measurement of human development is focused on 3 dimensions that are considered the most important for human life, namely longevity, knowledge and decent living standards according to UNDP. In general life opportunities are reflected in life expectancy (eo), knowledge is reflected in literacy rates (Lit) and average length of schooling (MYS) while decent living standards are reflected by purchasing power parity (PPP) . Each of the components is first calculated its index, so that a value between 0-100 with zero is the worst number and 100 is the best number. Simply put to calculate the index used the following formula:

$$Indeks\ X_i = \frac{[X_i - X_{i\,min}]}{[X_{i\,max} - X_{i\,min}]}$$

Where

X_i : indicator no-I; i=1 (life expectancy index), i = 2 (education index),

　　　and i = 3 (purchasing power index).

$X_{i\,max}$　: the maximum value of Xi.

Xi min: minimum value of Xi.

The maximum and minimum values in the formula above can be seen in the following table.

Table 2.1 Maximum Value and Minimum Value Xi

HDI Components (Xi)	Max Value	Minimum Value	Information
Life expectancy	85	25	UNDP Standards
Literacy numbers	100	0	UNDP Standards
Average length of school	15	0	UNDP uses a Combined Gross enrollment ratio
Purchasing power	737,720[a]	300.000 (1996) 360.000[b] (1999)	UNDP uses adjusted real per capita GDP

Source: Pemodelan Demand Penumpang Bandar Udara Domine Eduard Osok di Kota Sorong, Papua Barat, 2019

a) The maximum estimation at the end of PJP II is 2018.

b) Adjusting the old poverty line with the new poverty line.

All maximum and minimum values refer to the UNDP formula (1994) except for the value of the Real PPP adjusted to the situation of the State of Indonesia. For adjusted real per capita expenditure (real PPP) a maximum value of Rp.732,720.00 is taken from the real per capita expenditure for the city of Jakarta based on projected figures for 2018 assuming an increase of 6.5% over the 1993-2018 period. Whereas for the minimum value of real per capita expenditure used,

a constant of Rp.300,000.00 originating from twice the poverty line of South Sulawesi Province in rural areas in 1990. After the three index numbers above were generated, the global HDI value can be calculated, namely:

$$IPM = \frac{1}{3}\sum_{i=1} indeks\ X_i$$

Human Development index acts as X_2 for independent variable.

2.6.3 Education Index

The education index is a measure that shows the level of education of an area measured by 2 indicators, namely literacy rate and mean year school. Education Index acts as X_3 for independent variable.

2.6.4 Number of Labor Force

Number of labors acts as X_4 for independent variable. Number of labor force is the number of worker who are registered and work in all job. This can affect how labor exit or enter an area with regard to the large or small number of labor.

2.6.5 Number of Domestic Tourists

The number of domestic tourists is the number of visitors with tourist destinations, who come from within the country. The tourists tend to choose to travel by plane even though there are other alternatives such as ships but train is the most valuable mode to use to travel. Number of domestic tourists acts as X_5 for independent variable.

2.6.6 Number of Foreign Tourists

The number of foreign tourists is the number of visitors with tourist destinations coming from abroad. Sometimes they use the train to travel to each

city in Indonesia because the panoramic view be the best aspects. Number of foreign tourists acts as X_6 for independent variable.

2.6.7 Number of Student

Number of Students acts as X_7 for independent variable. Number of student is the number of student who are registered in all university and polytechnic. This can affect how student exit or enter an area with regard to the large or small number of student.

2.6.8 GRDP

Gross Regional Domestic Product or abbreviated as GRDP is the amount of value added goods and services resulting from all economic activities in all regions in a certain year or a certain period (usually in one year) with all aggregates calculated based on prices in the current year. GRDP in this modeling is stated in million rupiah. GRDP acts as X_8 for independent variable.

2.6.9 Population Purchasing Ability

The purchasing ability of the population is the ability of people to buy goods and services with the money they have. The level of people's purchasing power is influenced by several factors such as income, consumption expenditure, consumer price index, and expensiveness index. The purchasing power of the population is expressed in rupiah. Population purchasing ability acts as X_9 for independent variable.

2.6.10 Per Capita Income of the Population

Per capita income is the size of the average income of a population in an area. Per capita income is derived from regional income divided by the population

in the middle of the year. Per capita income is expressed in rupiah. Per capita income of the population acts as X_{10} for independent variable.

2.6.11 Per Capita Population Expenditure

Per capita expenditure is the average expenditure of the population in an area that is useful as a monitoring tool for the development of the standard of living of the population in an area. Expenditures per capita are expressed in rupiah. Per capita expenditures of the population acts as X_{11} for independent variable.

2.6.12 Statistical Test
2.6.12.1 Regression Analysis

Regression is the process of identifying mathematical equations (can be straight lines or curves that are more complex) that best matches the observational data. Identification is carried out on 2 variables namely the dependent variable (Y) and the independent variable (X) (Bell, MGH; Bonsall, PW; Leake, GR 1997). This method is the best method to analyze compare to another method. Linear regression models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

1. **Linear Regression Analysis**

   Linear regression analysis is intended to get the mathematical equation that is most suitable for the observational data, which in the process the dependent variable is only affected by 1 independent variable.

   Linear regression analysis has the general form of the equation:

   $$Y = a + bX$$

$$a = \frac{(\sum Y)\left(\sum Y^2\right) - (\sum X)(\sum XY)}{n \cdot \left(\sum X^2\right) - \left(\sum X^2\right)}$$

$$b = \frac{n \cdot (\sum X.Y) - (\sum X) \cdot (\sum Y)}{n \cdot \sum X^2 - (\sum X)^2}$$

Information:

Y: dependent variable

X: independent variable.

a: constant.

b: regression coefficient.

2. **Nonlinear Regression Analysis**

In addition to identification using linear equations, there are also identification of non-linear equations such as exponential equations, logarithms, hyperbole, polynomials, compounds, S functions, power functions, cubic functions and growth functions.

Some common forms of nonlinear regression analysis equations are as follows:

a. Exponential Function : $Y = a.b^X$ or $Y = a.e^{b.X}$

b. Logarithmic function : $Y = a + b.LnX$

c. Hyperbole function : $Y = a + b/X$

d. Polynomial function : $Y = a + b.X + c.X^2$

e. Compound function : $Y = a.b^X$

f. S function : $Y = e^{a+b/X}$

g. Power function : $Y = a.X^b$

h. Cubic function   : $Y = a + b.X + c.X^2 + d.X^3$

i. Growth function   : $Y = e^{a+bX}$

3. Multiple Linear Regression Analysis

Multiple linear regression analysis is intended to get the most suitable
mathematical equation to the observational data which in the process the
dependent variable is influenced by more than one variable together.
The general form of multiple linear regression equations is:

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_iX_i \; ; a = \frac{\sum Y}{n} \; b_1X_1 - b_2X_2$$

$$b_1 = \frac{\sum Y_2 i \,.\sum Y_2 i - \sum X_2 i \,.\sum X_2 iY_i}{n\,.\sum X_2{}^2 i - (\sum X_2 i)^2}$$

$$b_2 = \frac{n.\sum X_2 i \,.Y_i - \sum X_2 i \,.\sum Y_i}{n\,.\sum X_2{}^2 i - (\sum X_2 i)^2}$$

Information:

Y   : Dependent variable

X_1, X_2, X_3 : Independent variable

a   : a Constant

b_1, b_2, b_3 : Regression coefficient

2.6.13 Statistical Test Parameters

There are several statistical test parameters used in determining the best

model. These parameters are as follows:

1. Correlation coefficient (r)

Correlation is a method to determine the level of closeness of relationships between two or more variables, which is indicated by the value of the correlation coefficient. If the X value and the Y value represent the two variables being observed, the scatter plot illustrates the location points (X, Y) according to the coordinate system. If all points in the scatter plot appear to resemble a line as shown in Figure 2.1. Then the correlation is called linear. If Y tends to increase and X increases, as in Figure 2.1 on the left, then the correlation has a positive correlation or a direct correlation (+). Conversely, if Y tends to decrease while X increases, it looks like in Figure 2.1 to the right, then the correlation is called a negative correlation or an inverse correlation (-). Figure 2.2 shows the situation where there is no correlation between the two variables.
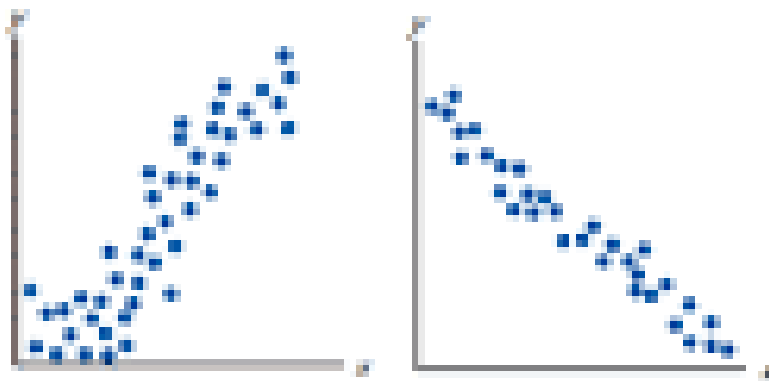
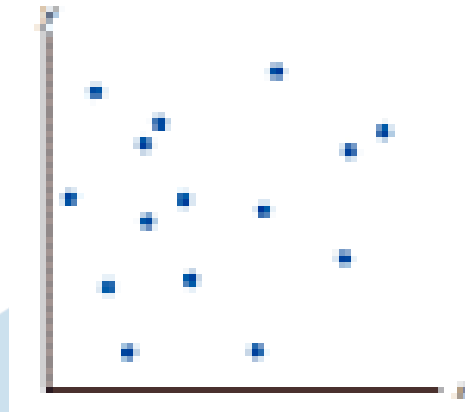**Figure 2.1 Correlation Scatter Diagram**

(Source: https://smartstat.wordpress.com/2010/11/21/korelasi-pearson/)

Figure 2.2 Scatter Diagrams No Correlation

(Source: https://smartstat.wordpress.com/2010/11/21/korelasi-pearson/)

The correlation between the two variables can be expressed by a
correlation coefficient (r). r values range from -1 to +1. Signs (+) and
signs (-) are used for each positive and negative correlation. Correlation
coefficient (r) can be found using the following formula:

$$r = \frac{\sum YX - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2) - (\sum X)^2 (n\sum Y^2 \ (\sum Y)^2)}}$$

For equations with I independent variables, the correlation coefficient (r)
is:

$$r = \sqrt{\frac{b1\sum x_1 y + \cdots + b_i \sum x_i y}{\sum y^2}}$$

Where:

Sxy $\quad : \sum XY - \frac{(\sum X)(\sum Y)}{N}$

Sx^2 $\quad : \sum X^2 - \frac{(\sum X)^2}{N}$

Sy^2 $\quad : \sum Y^2 - \frac{(\sum Y)^2}{N}$

2. **Coefficient of Determination (R^2).**

The coefficient of determination / R2 (R2) is calculated by multiplying the square of the correlation coefficient (r) is as follows:

$$r^2 = R^2 = \frac{[N \sum XY - (\sum X)(\sum X)]^2}{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}$$

3. **Standard Deviation**

Standard deviation or standard deviation is a measure of how wide the deviation / distribution of values from the average value. Standard deviation is a positive square root variant. Variants can be calculated with the following formula:

$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$; For sample

$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu_X)^2}{N}$; For population

Thus the standard deviation is

$S = \sqrt{S^2}$ and $\sigma = \sqrt{\sigma^2}$

The smaller the standard deviation of a model, the better the model.

4. **T-test.**

T-test is intended to test the independent variable (regression coefficient) whether it has an influence on the dependent variable. T value can be obtained by the following formula:

$$t = \frac{bi}{sbi}$$

Where:

t   : statistical testing for regression coefficients

bi   : regression coefficient

sbi   : standard deviation

5. **F-test.**

The F distribution test is intended to find out whether the variables that predict the formation of a regression meet the requirements, which are seen from the significant value of a certain level of confidence. This significant value is to compare the calculated F value with the F value of the table with a certain level of confidence. Said to be significant if the calculated F value is greater than the table F value. The amount of F value can be calculated by the following formula:

$$F = \frac{r^2(n - m - 1)}{m(1 - r^2)}$$

Where:

F   : F value of the regression line

n   : amount of data / sample

m   : number of independent variables

r^2 : correlation coefficient

2.6.14 All Possible Regression (APRE) Method

All possible regression (APRE) is one of the methods in the selection the best regression equation from several alternative regression equations available. In general, this method selects a set of estimator variables (X) best qualifies (for example those with the best R2) of several set the best regression equation according to the number of predicting variables (X). there are 3 the step of selecting

the best regression equation using the method

APRE, i.e.:

   a. Identifying all possible regression models, derived from all possible combinations of estimator variables (X).

   b. Choose the best regression model based on the number of variables expected (X) as an example two or three best regression models are chosen for each of the list of regression models is one variable X, two variables X, three variable X and so on.

   c. Evaluate several models that have been identified in the second step, so that it can be seen the effect of each variable that appears in the best regression model along with an increase in the value of R2 generated by each variable X. By considering this, the best equation can be generated.