# The Application of the Classical Assumption Test in Multiple Linear Regression Analysis (a Case Study of the Preparation of the Allometric Equations of Young Makila)

**Gun Mardiatmoko**
Department of Forestry, Pattimura University, Indonesia
g.mardiatmoko@faperta.unpatti.ac.id

## ABSTRACT

The preparation of allometric equations generally uses multiple linear regression equations. The use of this regression equation is usually not carried out through various tests but goes directly to the t test and F test stages. Therefore, this paper aims to provide an example of determining the regression equation through the classical assumption test stages. Makila (Litsea angulata BI) was chosen as an example because it provides environmental services, especially in dealing with climate change and has not been widely studied. The research method used is a quantitative method. Destructive sampling was taken and the material used consisted of 40 young Makila plants planted in polybags next to the greenhouse at the Faculty of Agriculture, Pattimura University, Ambon. The diameter of each sample was measured from the ground surface to a height of 5 cm and numbered 1 to 40. After being coded, the plant samples were cut down and all young trees were divided into stem segments, branches, twigs, leaves and roots. The root segments were collected carefully to facilitate their separation from the soil in polybags. The parts of the roots that still contain soil are cleaned with a machete and brush until they are free of soil and other dirt. Next, each segment was weighed for its wet weight (in g) and dry weight, thus enabling the determination of the biomass content (in g) for each segment. Data analysis was carried out using the Multiple Linear Regression Equation method. The regression model was examined for normality and suitability in predicting independent variables, ensuring there were no issues with multicollinearity, heteroscedasticity, and autocorrelation. The results of the normality test showed that the significance value for the remaining data was $0.813 > 0.05$, indicating normal distribution. TThe results of the multicollinearity test show that there is no multicollinearity problem, this is indicated by the VIF value of the two independent variables (tree diameter and tree height each 1.049) < 10 and the Tolerance value (tree diameter and tree height 0.953 each) > 0.100. The results of the heteroscedasticity test show that the two independent variables have a significance value of more than 0.05. The results of the autocorrelation test show the Durbin Watson (DW) value = 1.956 with a range of 1.65 – 2.35 which means there is no correlation problem. The results yielded a multiple linear regression: $Y = -1131.146 + 684.799X_1 + 4.276X_2$, where Y is biomass, $X_1$ is the diameter, and $X_2$ is the tree height. Based on the results of the t-test: variable $X_1$ partially affected Y while variable $X_2$ partially had no effect on Y. The F-test indicated that variables $X_1$ and $X_2$ jointly affected Y with R Square: 0.919 or 91.9% and the rest was influenced by other unexplored factors. To simplify biomass prediction and field measurement, a regression equation that used only 1 independent variable, namely tree diameter, was used for the experiment. Allometric equation only used 1 variable, $Y = -1,084,626 + 675,090 X_1$, where $X_1$ = tree diameter, Y = Total biomass with R = 0.957, and R square = 0.915. The regression equation for young Makila plant provides assurance that the regression equation obtained is accurate in estimation, unbiased and consistent. This is because in data processing classical assumptions have been tested. This equation can save time, costs and energy, as well as make measurements easier in the field so that future researchers can develop allometric equations in other plants for efficiency.

——————————— ◆ ———————————

## A.   INTRODUCTION

Makila (Litsea angulata BI) is an endemic wood species primarily found in the forest of Moluccas Provinces. The wood is highly valued for its natural durability and is extensively used mainly for residential house constructions, doors, furniture, and window products (Siruru & Titarsole, 2015). The Makila plant is locally known as 'kalangkala' in South Kalimantan, Central Indonesia, and is a species of the genus Litsea, belonging to the Lauraceae family. Most of these plants thrive in the wild in forest areas and are usually cultivated to extract their sturdy wood as building materials (Fitriyanti; et al., 2020). Other species such as Litsea boerlagei K. are spread in the Maluku region (Eastern Indonesia) and their synonym names include Litsea ambigua (Blume) Boerl. and Malapoenna ambigua (Blume) Kuntze (Hassler, 2023).

The genus Litsea is dominant in tropical and subtropical areas of India, China, Taiwan, and Japan. This plant has been widely explored due to its medicinal properties and is used traditionally to treat various gastrointestinal diseases, including diarrhea, stomach ache, indigestion, and gastroenteritis, as well as diabetes, colds, edema, asthma, arthritis, and traumatic injuries. Furthermore, Litsea is renowned for its essential oils, which have shown effectiveness against several bacteria, provide acute toxicity, possess antioxidant, antiparasitic, cytotoxic, genetic properties, and prevent several types of cancer (Kamle et al., 2019).

Currently, research has been carried out on the chemical composition, antioxidant, toxicity, and antimicrobial activity of *L. angulata* BI leaf essential oil in vitro. The results showed that the essential oil from *L. angulata* BI leaves was non-toxic, with the potential as an antimicrobial agent against oral pathogens (Putri, 2020). Apart from the leaves, the bark of *Litsea glutinosa*, (Lour.) C.B.Rob, a member of the Lauraceae family, is commonly found in the western region of Indonesia and can be used as a traditional medicine to treat rheumatism, diarrhea, and dysentery (Almurdani et al., 2023). In India, this tree serves as a binding agent in the incense industry in tablet formulations and as a plaster for broken limbs (Ramana & Raju, 2017). The wood is used locally and also traded as a source of '*medang*' (Fern, 2024).

Makila trees are often found in the Maluku forest and are also components of traditional agroforestry areas or '*dusung*' (Sahureka et al., 2021). This is because the hamlet area is usually filled with agricultural and forestry crops, thereby serving as a carbon store and playing an indirect role in controlling climate change. This is consistent with international efforts to carry out climate change mitigation by reducing emissions from forest land or Reducing Deforestation and Degradation Emissions (REDD+) (Goetz et al., 2015). Due to increased cooperation and a more comprehensive strategy to identifying deforestation and lowering emissions or global warming, REDD+ is able to reduce deforestation annually. Several regions are attempting to implement policy changes through REDD+ in order to further the goal of enhancing deforestation and ecology in their areas (Wahyuni & Suranto, 2021).

The determination of tree allometric equation is also an important tool for measuring forest resources. These equations are a statistical model primarily used to express the relationship between various components of a tree in relative size. Consequently, foresters transit from simple tree measurements such as stem diameter to characteristics including carbon stock and tree biomass, which are more difficult, expensive, and damaging to measure. This method considers the use and development of countries, biomes, climate, and species to minimize errors in propagation, increase accuracy, and reduce the bias resulting from the general model.

Regarding the categorization of allometric equations in the database, it is carried out on biomass, volume, carbon stock, and tree height. The biomass equation represents half of the total equation listed, followed by the volume. Meanwhile, the height diameter equation is included as part of the assessment information, which is often used as a reference when calculating tree volume and biomass (Kassim, A.R.;Sabri, M.D.M; Kamarudin, M.F.; Birigazzi, 2014).

The morphological conditions of *Litsea spp* trees including trunk, branching, and crown, particularly the relationship between trunk diameter and total tree height, trunk diameter and crown width, and total tree height and crown width, in East Kalimantan (Central Indonesia) Educational Forest have been investigated (Fajriansyah;Diana, R.; Matius, 2019). Based on the results, the regression coefficient values between diameter and height, as well as diameter and crown width showed moderate correlation values. This has led to extensive research on the Litsea genus concerning medicinal uses, construction, and morphological characteristics. There is no significant gap in understanding its relationship with biomass content, specifically through destructive sampling for handling climate change.

Forest biomass is calculated using a standard general allometric equation and can be implemented in large forest areas. Therefore, this research was carried out to establish an allometric equation to relate the diameter, height, and biomass of *L. angulata* BI, serving as a local model to support Reducing Emission from Deforestation and Forest Degradation (REDD+). Besides that, because in the application of multiple linear regression equations various tests are rarely carried out, but directly to the t test and F test stages, the classical assumption test stage is carried out.

Previous research conducted by Zaenal et al., (2020) stated that the biomass of mixed species trees in the study area could be estimated accurately using the M7 model which uses diameter, height and wood density variables. Other research states that by using allometric equations, the upper biomass, lower biomass and total biomass of the tree can be estimated only using the diameter parameter at breast height of Siregar (Wahyuni & Suranto, 2021). According to Fadillah (2023), tree biomass measurement estimates can use allometric equations to reduce destructive effects. This research is useful for saving time, costs and energy, as well as making it easier to measure in the field which is also related to practical research. His contribution to the scientific literature is adding scientific references related to the use of allometric equations in calculating biomass.

## B. METHODS
## 1. Study Site and Samples of young Makila plants

This research was carried out in 2 stages, which included the collection of data in the field and the laboratory testing of samples. The material used consisted of 40 young Makila vegetation (*L. angulata* BI) planted in polybags beside the greenhouse of the Faculty of Agriculture, Pattimura University, Ambon, as shown in Figure 1. The ages of young Makila plants were 1, 2, and 3 years, while the planting medium was red and yellow podzolic soil. This type of soil is characterized by low productivity due to intensive leaching and further weathering, resulting in poor soil physical and chemical properties (Abdillah & Aldi, 2020)**].**

Therefore, to enhance soil fertility and nutrient content, the media was enriched with NPK fertilizer in a 3:1 ratio, as shown in Figure 1.



**Figure 1.** Research location map

Maintenance of the Makila plant was carried out regularly every month to ensure the cleanliness of the planting location and remove lianas, grass, and other disruptive vegetation. Although there was no specific action for watering plants, it was only carried out incidentally, when extreme drought occurred due to a long drought. Plant diameter ranged from 1.5 cm to. 5.0 cm and the height varied from 0.65 m to 3.6 m. Samples of young Makila plants are presented in Figure 2.



**Figure 2.** Young plants of Makila (*L. angulata* BI)

## 2.   Field Data Collection Steps

The equipment used for collecting data in the field were hand saws, meters, scales, sieves, newsprints, machetes, tarpaulins, plastic bags, and stationery. Meanwhile, ovens and analytical scales were used for laboratory testing of samples. The diameter of each sample was measured from aboveground level to a height of 5 cm and numbered 1 to 40. After coding, the plant samples were cut down and all young trees were divided into segments for stems, branches, twigs, leaves, and roots. The root segments were collected carefully to facilitate each separation from the soil in the polybag. The part of the root that still contained soil was cleaned with a machete and brush until it was clean of soil and other debris. Subsequently, each segment was weighed for its wet (in g) and dry weight, allowing the determination of biomass content (in g) for each segment.

### 3. Data Processing

Data processing was carried out using the Multiple Linear Regression Equation method. The dependent variable (Y) was biomass and the 2 independent variables were diameter ($X_1$) and height ($X_2$) with the general formula:

$$\hat{y} = b0 + b1*x1 + \ldots + bn*xn + \epsilon \qquad (1)$$

Multiple linear regression makes all the same assumptions as simple linear regression, following the points stated below:

a. Homogeneity of Variance (Homoscedasticity): The magnitude of the error that occurs in the prediction does not change significantly at various values of the independent variable.

b. Independence of observations: In data collection, observations are obtained through statistically valid sampling methods, without hidden relationships between variables. Some of the independent variables also correlate with each other, necessitating the importance of checking before developing a regression model. When two independent variables are excessively highly correlated ($r2 > \sim 0.6$), only one should be used in the regression model.

c. Normality: The data collected is normally distributed.

d. Linearity: The line of best fit through the data points is a straight line, without a curve or grouping factor (Bevans, 2023).

Regression analysis constitutes a significant part of supervised machine learning, consisting of the prediction of a continuous independent target from a set of other predictor variables. The difference between binary classification and regression is in the target range. This is because, in binary classification, the target only consists of two values, usually encoded as 0 and 1), while in regression the target has multiple values (Chicco et al., 2021). Regression analysis plays a significant role in various scientific fields, as it seeks to establish a functional relationship from data to explain or predict a natural phenomenon based on other observable factors. Most regression analyses depend on the least squares method to estimate the parameters in the model. However, this method is usually formed with several assumptions, such as linearity, absence of autocorrelation, no multicollinearity, homoscedasticity, and normally distributed errors (Cahyandari, R.; Hisani, 2012). The use of multiple linear regression allows for the exploration of relationships between the dependent variable and two or more independent variables (Padilah, T.N; Adam, 2019). To facilitate this process, a correlation analysis is used to measure the level of relationship between two variables. Furthermore, correlation analysis explains how the regression line describes the variation of the dependent values and the strength of the relationship between the two variables (Chandra et al., 2022).

### 4. Data Analysis

Data analysis was carried out examining the availability of missing datasets through frequency analysis. This was followed by an assessment of multivariate normality, linearity, freedom from extreme values, and multilinear relationships with the assumption of multiple regression analysis. Multivariate normality analysis was carried out after the univariate normality assumption was completed for each quantitative variable, by examining the skewness coefficient and the histogram chart. Subsequently, a scatter diagram matrix was prepared for the assumption of multivariate linearity. Simple correlation test, variance increase factors (VIF), tolerance values, and Condition Index (CI) were carried out to determine the potential for double ties within the data. When the data analyzed met the requirements for the various assumption set, linear regression analysis was carried out (Uyanik, G.K;.Güler, 2013). The data collected from various analyses were processed using Microsoft Excel and the SPSS software package (ver.21) (Mondal & Mondal, 2016).

In the regression assumption test, the variance homogeneity test is used to determine whether the variances of two or more samples are considered the same or homogeneous. Independence of observations is carried out to find out whether there is a relationship between two categorical variables, in other words to find out the independence between row and column variables. The normality test is carried out to test whether in a regression model, an independent variable and a dependent variable or both have a normal or non-normal distribution. The linearity test is carried out to determine whether there is a linear relationship between the dependent variable and each independent variable to be tested (Ghozali, 2016).

### C. RESULT AND DISCUSSION

### 1. Classic Assumtion Test

The data collected from the 40 young Makila samples researched were processed, leading to the results of Descriptive Statistical Analysis, as presented in Table 1.

**Table 1**. Results of Statistical Descriptive Analysis

| | | Tree Diameter (cm) ($X_1$) | Tree Height (m) ($X_2$) | Bole (g) | Twig (g) | Leaf (g) | W above ground (g) | W below ground (g) | Total Biomass (g) (Y) |
|---|---|---|---|---|---|---|---|---|---|
| N | Valid | 40 | 40 | 40 | 40 | 37 | 40 | 40 | 40 |
| | Missing | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 |
| Mean | | 3.060 | 3.932 | 344.30 | 160.82 | 281.30 | 765.32 | 215.82 | 981.15 |
| Std. Error of Mean | | .1493 | 1.571 | 37.693 | 19.724 | 38.266 | 84.179 | 23.468 | 105.400 |
| Median | | 3.00 | 2.50 | 255.00 | 165.50 | 192.00 | 613.50 | 204.50 | 792.50 |
| Std. Deviation | | .944 | 9.935 | 238.393 | 124.745 | 232.760 | 532.396 | 148.422 | 666.607 |
| Variance | | .892 | 98.706 | 56,831.241 | 15,561.379 | 54,177.437 | 283,445.610 | 22,028.969 | 444,365.464 |
| Range | | 3.60 | 2.95 | 903 | 474 | 866 | 1,997 | 546 | 2,392 |
| Minimum | | 1.40 | .65 | 23 | 4 | 47 | 82 | 10 | 105 |
| Maximum | | 5.00 | 3.60 | 926 | 478 | 913 | 2079 | 556 | 2497 |
| Sum | | 122.40 | 92.93 | 13,772 | 6,433 | 10,408 | 30,613 | 8,633 | 39,246 |

a. Multiple modes exist. The smallest value is shown

Based on Table 1, the results obtained are that the Results of Statistical Descriptive Analysis with a sample size of 40 young Makila have a minimum value of 1.40 for Tree Diameter (cm) and 0.65 for Tre Height (m). Maximum value 5.00 for Tree Diameter (cm) and 3.60 for Tree Height (m). It was also found that all data was valid and there were no missing items. Based on the results of the Classical Assumption Test, the following results were obtained:

a. Residual Normality Test

This normality test was used to determine whether the residual value was normally distributed. Residual values that were normally distributed indicated a good regression model. The detection was carried out by observing the pattern of data distribution at the diagonal source on the standard P-P Normal Plot regression graph, which was the basis for decision-making. When the data spread out and followed a diagonal line, the regression model was declared normal and suitable for predicting the independent variable and vice versa. The results of the normality test were presented in the normal P-P Plot Chart Regression output, as shown in Figure 3.
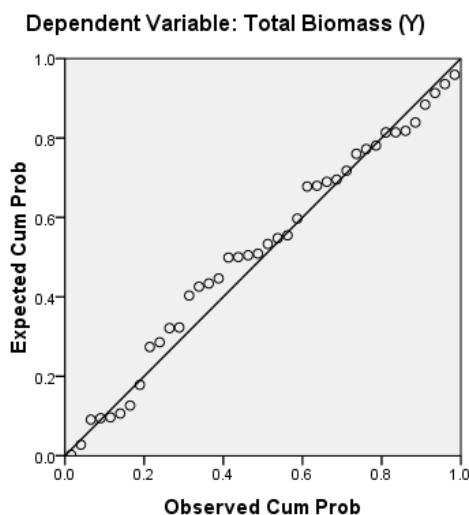


**Figure 3.** Normal P-P Plot of regression standardized

Based on Figure 3, Normal P-P Plot of regression standardized residuals indicated that the dots spread around the line and followed the diagonal line, confirming the normality of the regression model. This situation rendered the model suitable for predicting the independent variable and vice versa. To enhance the suitability of this assessment, the process can also be accompanied by another normality test, such as the One-Sample Kolmogorov Smirnov test method. The results of the residual normality test (NPar Test) are presented in Table 2.

The residual is the difference between the observed and predicted values in the regression model, which is calculated as Remaining = Observed value – Predicted value. When these observed values are plotted and overlay the fitted regression line, the residual for each observation becomes the vertical distance between the observation and the regression line. Moreover, the standardized residual is used to identify outliers in a regression model. In practice, standardized residuals with absolute values greater than 3 are often considered to be outliers. This does not mean that the observation will

be removed from the model but further investigation is required to verify whether these outliers are caused by data entry errors or unusual events (Zach, 2020).

**Table 2.** Residual Normality Test Results

**One-Sample Kolmogorov-Smirnov Test**

| | | Tree Diameter (cm) ($X_1$) | Tree Height (m) ($X_2$) | Total Biomass (g) (Y) | Unstandardized Residual |
|---|---|---|---|---|---|
| N | | 40 | 40 | 40 | 40 |
| Normal Parameters[a] | Mean | 3.0600 | 3.9320 | 981.15 | 0.0000000 |
| | Std. Deviation | .94457 | 9.93508 | 666.607 | 1.89803389E2 |
| Most Extreme Differences | Absolute | 0.101 | 0.468 | 0.134 | 0.101 |
| | Positive | 0.065 | 0.468 | 0.134 | 0.055 |
| | Negative | -.101 | -.384 | -.113 | -.101 |
| Kolmogorov-Smirnov Z | | 0.639 | 2.962 | 0.847 | 0.637 |
| Asymp. Sig. (2-tailed) | | 0.809 | 0.000 | 0.470 | 0.813 |

a. Test distribution is Normal.

Based on Table 2, the significance value (Asym.sig 2 tailed) for the data residual is 0.813 > 0.05, indicating normally distributed. This condition further confirms that the regression model is suitable for predicting independent variables and vice versa.

b. Multicollinearity Test

Multicollinearity is a situation where there is a perfect or close linear relationship between independent variables in a regression model. This situation arises due to a perfect linear function for some or all of the independent variables. Consequently, the effects of independent and dependent variables are difficult to observe. Symptoms of multicollinearity can be obtained by observing VIF and Tolerance values. When the VIF value < 10 and Tolerance > 0.1, it is stated that there is no multicollinearity. The results of the multicollinearity test are presented in Table 3.

**Table 3**. Multicollinearity Test Results

| Model | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Standard Error | Beta | t | Sig. | Tolerance | VIF |
| 1 (Constant) | -1,131.146 | 111.322 | | -10.161 | 0.000 | | |
| Tree Diameter (cm) ($X_1$) | 684.799 | 33.832 | 0.970 | 20.241 | 0.000 | 0.953 | 1.049 |
| Tree Height (m) ($X_2$) | 4.276 | 3.217 | 0.064 | 1.329 | 0.192 | 0.953 | 1.049 |

a. Dependent Variable: Total Biomass (Y)

The results presented in Table 3 showed that there was no multicollinearity problem, as indicated by the VIF value for the two independent variables (tree diameter and tree

height each 1.049) < 10 and the Tolerance value (tree diameter and tree height each 0.953) > 0.100.

c. Heteroscedasticity Test

Heteroscedasticity is a situation where there are differences in residual values for all observations in a regression model. This test method is carried out using the Glejser Test, by regressing the independent variable on the absolute residual value. The residual is the difference between the observed and predicted values of the variable Y, and is absolute, indicating that all values are positive. When the significance value between the independent variables and absolute residuals is > 0.05, there is no heteroscedasticity problem. The results of this regression, specifically the Glejser Heteroscedasticity Test, are presented in Table 4.

**Table 4.** Heteroscedasticity Test Results

| | Coefficients[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Unstandardized Coefficients | | Standardized Coefficients | | | | Collinearity Statistics | |
| Model | B | Standard Error | Beta | t | Sig. | Tolerance | VIF |
| 1 (Constant) | -1,131.146 | 111.322 | | -10.161 | 0.000 | | |
| Tree Diameter (cm) $(X_1)$ | 684.799 | 33.832 | 0.970 | 20.241 | 0.000 | 0.953 | 1.049 |
| Tree Height (m) $(X_2)$ | 4.276 | 3.217 | 0.064 | 1.329 | 0.192 | 0.953 | 1.049 |

a. Dependent Variable: Total Biomass (Y)

The results of the heteroscedasticity test in Table 4 show that the two independent variables (tree diameter and tree height) have a significance value of more than 0.05. This indicated that there was no problem heteroscedasticity in this regression model.

d. Autocorrelation test

Autocorrelation is a condition where there is a correlation between residuals in period t with the previous period (t-1) in a regression model. A good regression model should not have any autocorrelation problem. The results of the autocorrelation test obtained are presented in Table 5.

**Table 5.** Autocorrelation Test Results

| Model | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Standard Error | Beta | t | Sig. |
| 1 (Constant) | 67.273 | 67.897 | | 0.991 | 0.328 |
| Tree Diameter (cm) $(X_1)$ | 27.463 | 20.635 | 0.215 | 1.331 | 0.191 |
| Tree Height (m) $(X_2)$ | -1.667 | 1.962 | -.137 | -.850 | 0.401 |

a. Dependent Variable: ABS_RES

The results of the autocorrelation test presented in **Table 5** show that the Durbin Watson (DW) value = 1.956, ranging from 1.65 – 2.35. This condition indicated that there is no autocorrelation problem.

## 2. Result of Multiple Linear Regression Analysis

Multiple linear regression analysis was used to determine the effect of independent variables on the dependent variable, both partially with the t-test and collectively through the F-test. The multiple linear regression analysis was used to formulate the regression equation and discover the increasing or decreasing value of the variable Y for the change in X. The results obtained after the data were processed with Microsoft Excel and SPSS were presented in Table 6.

**Table 6.** Processing Results of Multiple Linear Regression Analysis

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-----|----------|-------------------|----------------------------|---------------|
| 1 | 0.959[a] | 0.919 | 0.915 | 194.866 | 1.956 |

a. Predictors: (Constant), Tree Height ($X_2$), Tree Diameter ($X_1$)
b. Dependent Variable: Total Biomass (Y)

Based on Table 6, the regression equation is obtained, namely: $Y = -1131.146 + 684.799X_1 + 4.276X_2$: (1) Constant -1,131.146, indicating that when $X_1$ and $X_2$ are 0, the magnitude of Y is negative -1,131.146; (2) The variable regression coefficient $X_1$ = 684.799, indicating that every time $X_1$ increases by 1 unit, it will increase Y = 684,799 units, assuming the values of the other independent variables remain constant; and (3) The variable regression coefficient $X_2$ = 4.276, indicating that with every increase in $X_2$ by 1 unit, it will increase Y = 4.276 units, assuming the other independent variables have a fixed value.

a. t-test

t-test in multiple regression was used to determine whether the independent variable regression model had a partially significant effect on the dependent variable. The results of the t-test presented in Table 6 showed that Variable $X_1$ partially influenced Y, with a significance value of <0.05 (0.000 <0.05). This effect was positive due to the coefficient, indicating that an increase in $X_1$ led to a rise in Y. Meanwhile, variable $X_2$ partially did not affect Y because the significance value was > 0.05 (0.192 > 0.05). For more details, the Histogram and Scatterplot are presented in Figures 4 and Figure 5.
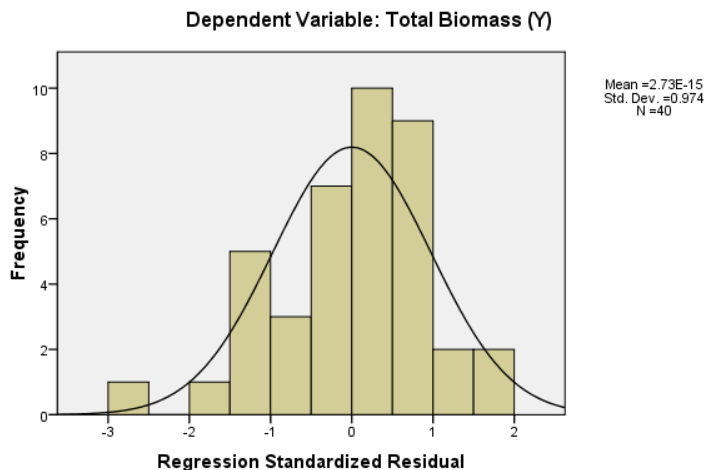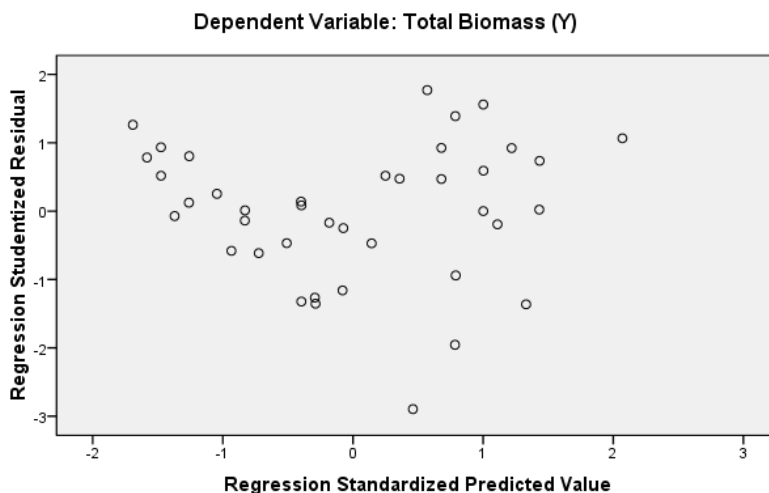
**Figure 4.** Histogram



**Figure 5.** Scatterplot

b. F test: The results of the F test with analysis of variance are presented in Table 7.

In Table 7, a significance value < 0.05 (0.000 < 0.05) was obtained, indicating that variables $X_1$ and $X_2$ collectively affected Y. The results of the regression and determination analysis (R Square) are presented in Table 8.

**Table 7.** F Test Results with Analysis of Variance

| No | Model | Sum of Squares | df | Mean Square | F | Sig. |
|----|-------|----------------|-----|-------------|-----|------|
| 1 | Regression | 1.593E7 | 2 | 7,962,632.686 | 209.694 | 0.000a |
| 2 | Residual | 1,404,987.728 | 37 | 37,972.641 | | |
| | Total | 1.733E7 | 39 | | | |

a. Predictors: (Constant), Tree Height (m) ($X_2$), Tree Diameter (cm) ($X_1$)
b. Dependent Variable: Total Biomass (g) (Y)

**Table 8**. Results of Regression Analysis and Determination (R Square)

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-----|----------|-------------------|----------------------------|---------------|
| 1 | 0.959a | 0.919 | 0.915 | 194.866 | 1.956 |

a. Predictors: (Constant), Tree Height ($X_2$), Tree Diameter ($X_1$)

Table 8 showed that the regression coefficient value R was 0.959 or close to 1, indicating a very strong relationship between tree diameter ($X_1$), tree height ($X_2$), and total biomass (Y).

An $R^2$ of 0.919 was obtained, indicating that the variables $X_1$ and $R^2$ had been used over several decades to test the goodness of fit of regression models. However, it was under criticism over the last three to four decades due to diverse limitations and appropriateness for applicability to nonlinear models. This research reviewed some of the significant investigations, regarding the strengths and weaknesses of $R^2$, and cautioned the application of these statistics in the predictive methodology.

According to (Sapra, 2014), $R^2$ has been used for decades to test the suitability of regression models, making it a useful statistic in predictive modeling in the future. Most users remained emotionally attached to these statistics for a long time due to their ability to evaluate the performance of their model. However, there were negative comments, emphasizing the need for caution when applying $R^2$. This indicated that the strengths and weaknesses must be assessed first before implementing $R^2$.

The notation for correlation and regression coefficients needs to be established. This is because using r or R, but not $r^2$ or $R^2$, is appropriate to avoid confusion with other coefficients that have different meanings. As explained in **Table 6**, the equation model for estimating above and below-ground biomass of young Makila trees *(L.angulata* BI) was presented as follows $Y = -1131.146 + 684.799X_1 + 4.276X_2$, with an $R^2$ of 0.919. This supported previous research for young rain trees (*Albizia saman* (Jacq.) Merr) $Y = -10,310.50 + 1,820.89X_1 + 10.89X_2$, with $R^2$ was 0.859 (Mardiatmoko, 2016). Similarly, for young Canarium trees (*Canarium indicum* L.), the equation was $Y = -700,200 + 209.149X_1 + 3.922X_2$ with $R^2$ 0.386 (Mardiatmoko, 2022).

This research uses 2 variables, namely diameter and height plant, but there are also other variables such as wood density. According to (Thippawan et al., 2023), the best model for predicting biomass uses three stem variables: Diameter, Height, and Wood Density. These various variables are effective in predicting seed biomass because they are directly proportional to the accumulation of biomass in the seed stem. The use of the Wood Density variable requires destructive sampling of seedlings therefore it can be proposed without the Wood Density variable as an alternative model. The model that only uses Diameter and Height gives slightly higher Akaike information criterion (AIC) values, but the RMSE and Avg%ER values are almost the same, when compared with the best model with these three variables. AIC is a mathematical method for evaluating how well a model fits the data it produces and in statistics it is used to compare various possible models and determine which model best fits the data (Bevans, 2020).

Based on the t-test in multiple regression in Table 6, Variable $X_1$ (tree diameter) partially influenced Y (total biomass) but Variable $X_2$ (tree height) partially did not influence Y (total biomass). In this situation, considering a regression equation that uses only 1 independent variable, namely tree diameter is important to facilitate easy measurements in the field, and save time, costs, as well as energy. This is because measuring tree height in the field experiences many difficulties, particularly in a dense forest. Based on the results of simple linear regression

analysis, the equation obtained was Y = - 1,084,626 + 675,090 $X_1$, where $X_1$ = tree diameter and Y = total biomass, as presented in Table 9. Furthermore, there was also a very close relationship between tree diameter and total biomass with R = 0.957 and $R^2$= 0.915, as illustrated in Table 10. The data distribution and the regression line equation are presented in Figure 6.
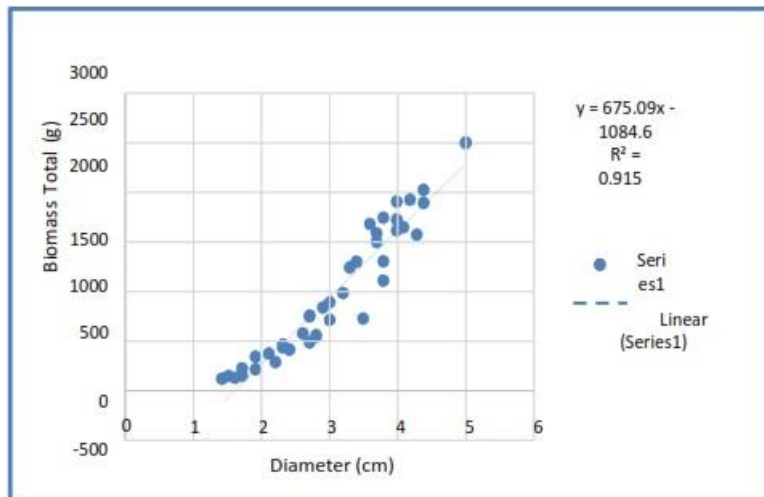


**Figure 6.** Data distribution and regression line equation

**Table 9.** Processing Results of Simple Linear Regression Analysis

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Coefficients[a]** | | | | | | |
| | **Unstandardized Coefficients** | | **Standardized Coefficients** | | | **Collinearity Statistics** | |
| **Model** | **B** | **Standard Error** | **Beta** | **t** | **Sig.** | **Tolerance** | **VIF** |
| 1 (Constant) | -1,084.626 | 106.739 | | -10.161 | 0.000 | | |
| Tree Diameter (cm) ($X_1$) | 675.090 | 33.366 | .957 | 20.233 | 0.000 | 1.000 | 1.000 |
| a. Dependent Variable: Total Biomass (g) (Y) | | | | | | | |

Based on Table 9, shows that by comparing the sig value of 0.000, which means it is not greater than 0.05. So it can be concluded that the data is not normally distributed. The data needs to be made normal first.

**Table 10.** Processing Results of Simple Linear Regression Analysis

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | 0.957[a] | 0.915 | 0.913 | 196.823 | 1.980 |
| a. Predictors: (Constant), Tree Diameter (cm) ($X_1$) | | | | | |
| b. Dependent Variable: Total Biomass (g) (Y) | | | | | |

Based on Table 10, it is known that the coefficient of determination value is 0.915. The coefficient of determination (R square) is 91.5%. This means that tree diameters influence total biomass and the rest is influenced by other variables not studied.

The research results of (Karyati; Ipor, Isa B; Jusoh, Ismail; Wasli, 2021) in Sarawak, Malaysia show that Above ground biomass from the genus Litsea is: L. elliptica Blume, L. costalis (Nees) Kosterm. var. nidularis Gamble, L. crassifolia Boerl, respectively 491 g, 73 g, 33 g medium for L. angulate, BI medium in this study 765.32 g. Above ground biomass of L. angulate BI is greater because the plant samples used range in diameter from 1.5 cm to. 5.0 cm while in plants in Sarawak it only ranges from 0.2 to 4.8 cm. Thus, the AGB results obtained are different even though they come from the same genus or even for the same species, but if the habitat or environmental ecosystem is different, the AGB will be different too. In this way, allometric equations applied to the same species but in different locations and habitats will result in different allometric equations. We recommend increasing research on the allometric equations of young plants at seedling and sapling levels, especially in tropical forests with many types. The variables used are not only diameter, height, wood density but can be other variables, but the most important thing is to process the data first using the classical assumption test.

## D.    CONCLUSION AND SUGGESTIONS

This research obtained an allometric equation for young Makila plant (L.angulata BI) of $Y = -1131.146 + 684.799X_1 + 4.276X_2$, where Y indicated biomass, X1 represented diameter, and X2 was the plant height with R2 0.919. Before this allometric equation was established, a classical assumption test was carried out in data processing. Based on the results of the t-test: variable X1 partially affected Y, while variable X2 partially did not affect Y. The F-test analysis showed that variables X1 and the rest were influenced by other unexplored factors.

In summary, the allometric model describes the relationship between the size or growth of a single part and the entire structure of a living organism. It is a useful tool for explaining systematic changes. The allometric equation model makes it easy to estimate a tree's biomass by just entering its diameter, height, or a combination of the two, which allows for the calculation of the ecosystem's standing biomass. In this situation, a regression equation that used only 1 independent variable, namely tree diameter was used to save time, costs, as well as energy, and facilitate easy measurements in the field. This was because measuring tree height in the field experienced many difficulties, particularly in a dense forest. Allometric equation only used 1 variable, $Y = -1,084,626 + 675,090 X_1$, where X1 = tree diameter, Y = total biomass, R = 0.957, and R2 = 0.915. Considering the potential for time, costs, and energy, as well as ease of measurement in the field, it would be better to predict the biomass of young Makila plant by measuring the tree diameter and avoiding height. The regression equation for young Makila plant provides assurance that the regression equation obtained is accurate in estimation, unbiased and consistent. This is because in data processing classical assumptions have been tested. In the next step, authors or researchers can apply allometric equations to other plant types for ease of implementation in the field.

## REFERENCES

Abdillah, M. H. ., & Aldi, M. (2020). Aplikasi Limbah Padat Karet Remah Pada Tanah Podsolik Merah Kuning Terhadap Ketersediaan Hara Makro Dan Perbaikan Sifat Fisika Tanah. *EnviroScienteae*, *16*(2), 264–275. https://doi.org/10.20527/es.v16i2.9658

Almurdani, M., Hendra, R., & Teruna, H. Y. (2023). Antioxidant activities of medang lendir (Litsea

glutinosa) stem bark. *Pharmacy Education*, *23*(2), 122–125. https://doi.org/10.46542/pe.2023.232.122125

Bevans, R. (2020). *Akaike Information Criterion | When & How to Use It ( Example )*. https://www.scribbr.com/statistics/akaike-information-criterion/

Bevans, R. (2023). *Multiple Linear Regression | A Quick Guide ( Examples )*. Scribbr. https://www.scribbr.com/statistics/multiple-linear-regression/

Cahyandari, R.; Hisani, N. (2012). Model Regresi Linier Berganda Menggunakan Penaksir Parameter Regresi Robust M-Estimator (Studi Kasus: Produksi Padi di Provinsi Jawa Barat Tahun 2009). *Istek*, *VI*(1), 85–92. https://journal.uinsgd.ac.id/index.php/istek/article/view/291

Chandra, Y. I., Riastuti, M., Tinggi, S., Informatika, M., & Sti, J. (2022). Aplikasi Perhitungan Statistik Koefisien Regresi Korelasi Linier Untuk Produksi Minyak dan Gas Bumi. *Jurnal Esensi Infokom*, *6*(2), 19–25. https://doi.org/10.55886/infokom.v6i2.500

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*(1), e623. https://doi.org/10.7717/peerj-cs.623

Fadillah, M. (2023). *Estimasi biomassa dan cadangan karbon pohon di kawasan konservasi Taman Wisata Alam (TWA) Iboih, Kota Sabang*. Universitas Islam Negeri Ar-Raniry. https://repository.ar-raniry.ac.id/id/eprint/32994/

Fajriansyah;Diana, R.; Matius, P. (2019). Profil Pohon pada jenis Litsea spp . di Hutan Pendidikan. *Makila*, *13*(2), 85–100. https://doi.org/10.30598/makila.v13i2.2435

Fern, K. (2024). *Tropical Plants Database*. Australian National Botanic Gardens. https://ntbg.org/database/plants/

Fitriyanti;, Qalbiah, S. ., & Sayakti, P. . (2020). Identifikasi Kulit Batang Kalangkala (Litsea Angulata Bi) Secara Makroskopik, Mikroskopik, Dan Skrining Fitokimia. *Parapemikir*, *9*(2), 1–9. https://doi.org/10.30591/pjif.v

Ghozali, I. (2016). *Application of multivariete analysis with IBM SPSS 23 program*. Semarang: Diponegoro University Publishing Agency.

Goetz, S. J., Hansen, M., Houghton, R. A., Walker, W., Laporte, N., & Busch, J. (2015). Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD+. *Environmental Research Letters*, *10*(12), 123001. https://doi.org/10.1088/1748-9326/10/12/123001

Hassler, M. (2023). *World Plants. Synonymic Checklist and Distribution of the World Flora. Version 16.3*. https://doi.org/10.48580/dfvll-3dd

Kamle, M., Mahato, D. K., Lee, K. E., Bajpai, V. K., Gajurel, P. R., Gu, K. S., & Kumar, P. (2019). Ethnopharmacological Properties and Medicinal Uses of Litsea cubeba. *Plants*, *8*(150), 1–13. https://doi.org/10.3390/plants8060150

Karyati; Ipor, Isa B; Jusoh, Ismail; Wasli, M. E. (2021). Allometric Equations to Estimate the Aboveground Biomass of Seedling and Sapling Plants in 10 and 20 Years Old of Secondary Forests in Sarawak , Malaysia. *Proceedings of the Joint Symposium on Tropical Studies (JSTS-19)*, *11*(1), 216–224. https://doi.org/10.2991/absr.k.210408.036

Kassim, A.R.;Sabri, M.D.M; Kamarudin, M.F.; Birigazzi, L. (2014). *Inventory of Tree Biomass and Volume Allometric Equations in Southeast Asia* (Issue June). UN-REDD PROGRAMME. https://www.un-redd.org/document-library/inventory-volume-and-biomass-tree-allometric-equations-southeast-asia-2018-update

Mardiatmoko, G. (2016). Allometric Equations For Predicting Above And Below-Ground Biomass Of Young Rain Tree [Albizia Saman (JACQ.) Merr.] To Handle Climate Change. *Asian Journal of Microbiology , Biotechnology & Environmental Sciences*, *18*(4), 821–830. http://www.envirobiotechjournals.com/article_abstract.php?aid=7365&iid=221&jid=1

Mardiatmoko, G. (2022). Allometric Equations For Predicting Biomass Of Young Canarium Tree ( Canarium indicum L .) To Handle Climate Change. *AMA*, *53*(01), 5471–5482. https://www.shin-norinco.com/article/allometric-equations-for-predicting-biomass-of-young-canarium-tree-canarium-indicum-l-to-handle-climate-change

Mondal, H., & Mondal, S. (2016). Sample size calculation to data analysis of a correlation study in Microsoft Excel®: A hands-on guide with example. *Int J Clin Exp Physiol*, *3*(4), 180–189.

https://ijcep.org/index.php/ijcep/article/view/288

Padilah, T.N; Adam, R. . (2019). Analisis regresi linier berganda dalam estimasi produktivitas tanaman padi di kabupaten karawang 1,2). *Fibonacci*, *5*(2), 117–128. https://doi.org/10.24853/fbc.5.2.117-128

Putri, A. S. (2020). Toxicity, antioxidant ability and inhibition of oral pathogens by monoterpene-rich essential oil of Litsea angulata Blume. *AGRICULTURE AND NATURAL RESOURCES*, *54*(1), 1–6. https://li01.tci-thaijo.org/index.php/anres/article/view/242192

Ramana, K. V., & Raju, A. J. S. (2017). Traditional and commercial uses of Litsea glutinosa ( Lour .) C . B . Robinson ( Lauraceae ). *Journal of Medicinal Plants Studies*, *5*(3), 89–91. https://www.semanticscholar.org/paper/Traditional-and-commercial-uses-of-Litsea-glutinosa-Ramana-Raju/3343744d57935c2511beab22bd712a8d6cbe91ce

Sahureka, M., Hadijah, M., & Marasabessy, H. (2021). Contribution of agrosilvopastoral pattern to sustainable communities economic improvement (study in Namaa Hamlet, Pelauw Village, Haruku Island Sub-District, Central Maluku Regency). *IOP Conference Series: Earth and Environmental Science*, *883*(1), 12071. https://doi.org/10.1088/1755-1315/883/1/012071

Sapra, R. L. (2014). Using R 2 with caution. *Current Medicine Research and Practice*, *4*(3), 130–134. https://doi.org/10.1016/j.cmrp.2014.06.002

Siruru, H. ., & Titarsole, J. (2015). Kandungan dan komponen kimia kayu makila pada arah aksial. *Makila*, *IX*(1), 52–62. https://docplayer.info/72167774-Kandungan-dan-komponen-kimia-kayu-makila.html

Thippawan, S., Chowtiwuttakorn, K., & Pongpattananurak, N. (2023). Allometric Models to Estimate the Biomass of Tree Seedlings from Dry Evergreen Forest in Thailand. *Forests*, *14*(1), 725. https://doi.org/10.3390/f14040725

Uyanik, G.K;.Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*(1), 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Wahyuni, H., & Suranto, S. (2021). Dampak deforestasi hutan skala besar terhadap pemanasan global di Indonesia. *JIIP: Jurnal Ilmiah Ilmu Pemerintahan*, *6*(1), 148–162. https://doi.org/10.14710/jiip.v6i1.10083

Zach. (2020). *What Are Standardized Residuals ?* https://www.statology.org/standardized-residuals/

Zaenal, M. S., Tiryana, T., & Muhdin, M. (2020). Allometric models for estimating tree biomass of dryland secondary forest in East Halmahera. *Jurnal Wasian*, *7*(2), 87–101. http://ejournal.forda-mof.org/ejournal-litbang./index.php/JWAS/article/view/5948