

# Week 4 Lab Report

Nguyen Tien Anh Quach

2024-02-15

## Mean tree diameter vs. Mean height

1.  $H_0$ : There is no significant linear relationship between mean height and mean tree diameter.

$H_a$ : There is a significant linear relationship between mean height and mean tree diameter.

2. Linear model result:

Mean tree height in forest plots has a significantly positive linear relationship with mean tree diameter (Fig. 1). With a 1 m increase in mean tree height resulting in a 1.86 cm increase in mean tree diameter ( $R^2 = 0.95$ , Adj.  $R^2 = 0.95$ ,  $F(1,71) = 1286$ ,  $p < 0.001$ ). The regression equation is :

$$\hat{y}_i = -8.08 + 1.86 \times mH.m_i$$

One thing to notice is that the intercept of the linear model is -8.08, which means that when the mean tree diameter is 0, the mean height of the tree is -8.08 m. This is not realistic!

3. Checking assumptions:

### a. Linearity

Residual plot showed that the residuals of most data points are randomly and closely scattered around the center line (Fig. 2). Linearity assumption is met!

### b. Normality

QQ plot showed that most data points are randomly and closely scattered around the normality line (Fig. 3). Therefore, normality assumption is met!

### c. Constant variance

The test for non-constant variance reported an insignificant result ( $p = 0.78$ ). Therefore, we failed to reject the null hypothesis and the assumption of homoscedasticity is met!

### d. Independence

The Durbin-Watson reported an insignificant result ( $p = 0.58$ ). Therefore, we failed to reject the null hypothesis and the assumption of independence is met!

## Mean tree height vs. Mean wood density

1.  $H_0$ : There is no significant linear relationship between mean height and mean wood density.

$H_a$ : There is a significant linear relationship between mean height and mean wood density.

2. Linear model results:

Mean wood density in forest plots has a significantly positive linear relationship with mean tree height (Fig. 4). With a 1 g/m<sup>3</sup> increase in mean wood density resulting in a 11.24 m increase in mean tree diameter ( $R^2 = 0.23$ , Adj.  $R^2 = 0.24$ ,  $F(1,71) = 22.27$ ,  $p < 0.001$ ). The regression equation is :

$$\hat{y}_i = 10.11 + 11.24 \times mWD.g.m3_i$$

One thing to notice and check later on is the low R-squared value, which is only 0.24. This could be due to outliers and/or high residuals.

### 3. Checking assumptions:

#### a. Linearity

Residual plot showed that the residuals of most data points are randomly and closely scattered around the center line (Fig. 5). However, point 60 is potentially the extreme outlier.

#### b. Normality

QQ plot showed that most data points are randomly and closely scattered around the normality line (Fig. 6). Several outliers are shown. I tested the normality of the residuals and p-value was reported to be significant, which was 0.29. Therefore, normality assumption is met!

#### c. Constant variance

The test for non-constant variance reported an significant result ( $p = 0.02$ ). Therefore, we rejected the null hypothesis and the assumption of homoscedasticity is NOT met!

#### d. Independence

The Durbin-Watson reported an insignificant result ( $p = 0.59$ ). Therefore, we failed to reject the null hypothesis and the assumption of independence is met!

As the constant variance assumption is NOT met, there may be outliers that need to be removed.

#### Detecting outliers:

The original dataset had 5 outliers, among which two are extreme outliers. I removed the extreme outliers and conducted the regression analysis again.

Mean wood density in forest plots still has a significantly positive linear relationship with mean tree height. With a  $1 \text{ g/m}^3$  increase in mean wood density resulting in a 6.60 m increase in mean tree diameter ( $R^2 = 0.07$ , Adj.  $R^2 = 0.06$ ,  $F(1,69) = 5.319$ ,  $p = 0.02$ ). The regression equation is :

$$\hat{y}_i = 12.86 + 6.60 \times mWD.g.m3_i$$

I also checked the assumptions, which are now all met. P-values are listed below:

1. Normality:  $p = 0.31$ .
2. Constant variance:  $p = 0.86$ .
3. Independence:  $p = 0.42$ .
4. Linearity: QQplot looks normal (Fig. 7).

*#Normality comment* Against better judgment, in the past we have used the `shapiro.test()` to assess normality. Remember that no test will show that your data has a normal distribution. Normality statistics show when your data is sufficiently inconsistent with a normal distribution that you would reject the null hypothesis of “no difference from a normal distribution”. However, when the sample size is small, even big departures from normality are not detected, and when the sample size is large, even the smallest deviation from normality will lead to a rejected null. In other words, if we have enough data to fail a normality test, we always will because real-world data won’t be clean enough. See (<http://www.r-bloggers.com/normality-and-testing-for-normality/>) for an example with simulated data. So, where does that leave us? Explore your data for large deviations from normality and make sure to assess heteroscedasticity and outliers. But, don’t get hung up on whether your data are normally distributed or not. As the author of the above link suggests: “When evaluating and summarizing data, rely mainly on your brain and use statistics to catch really big errors in judgment.”

## Appendix

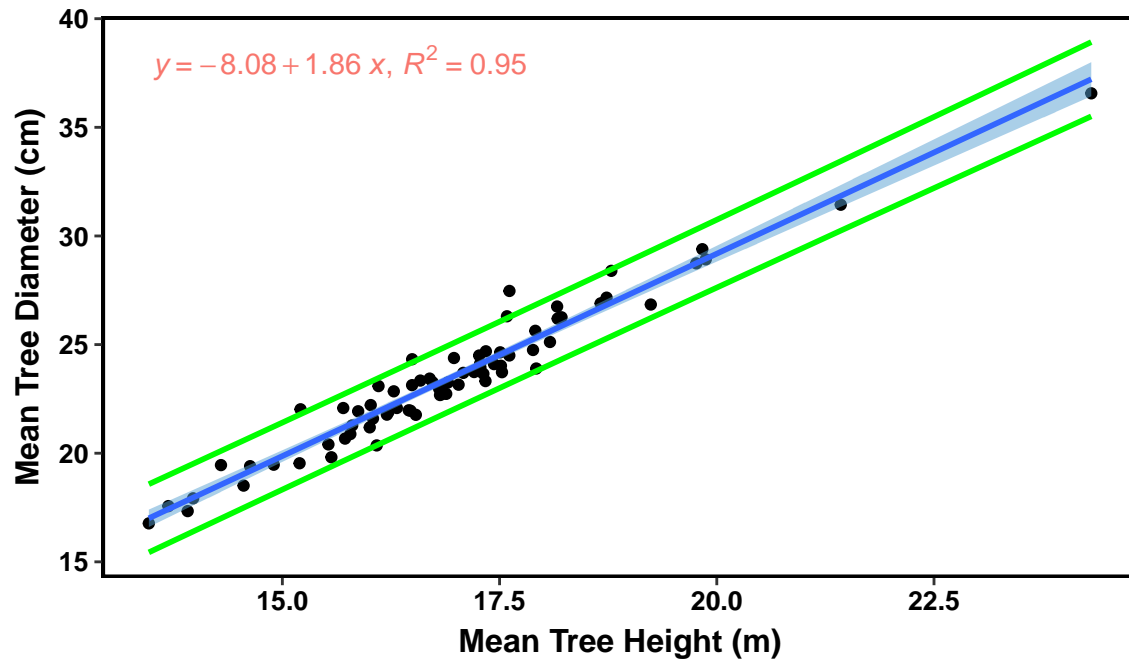


Figure 1. Linear regression of mean tree height and mean tree diameter with confidence and prediction intervals.

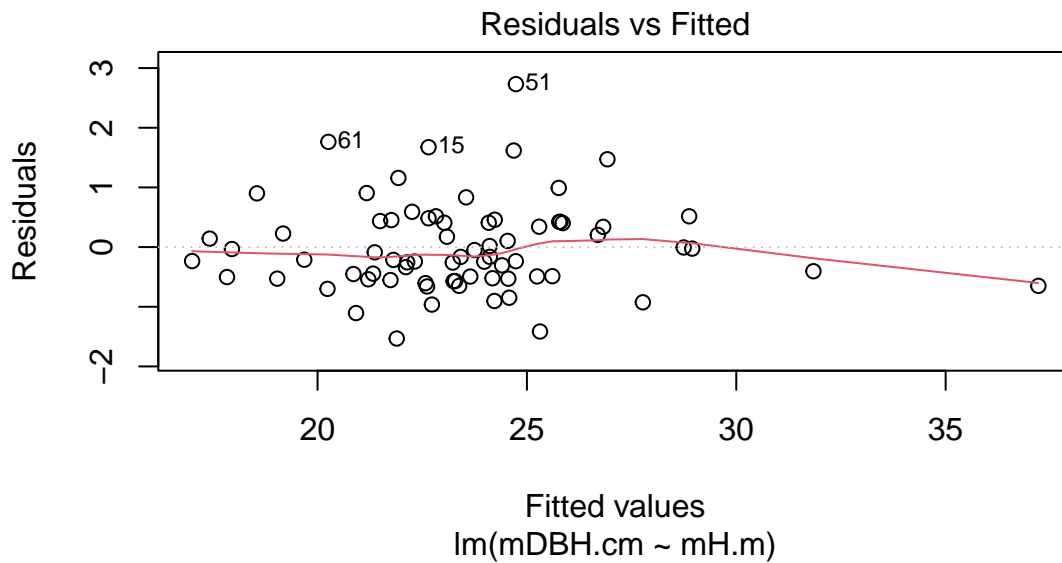


Figure 2. Residuals vs. Fitted values plot of the linear model between mean tree height and mean tree diameter.

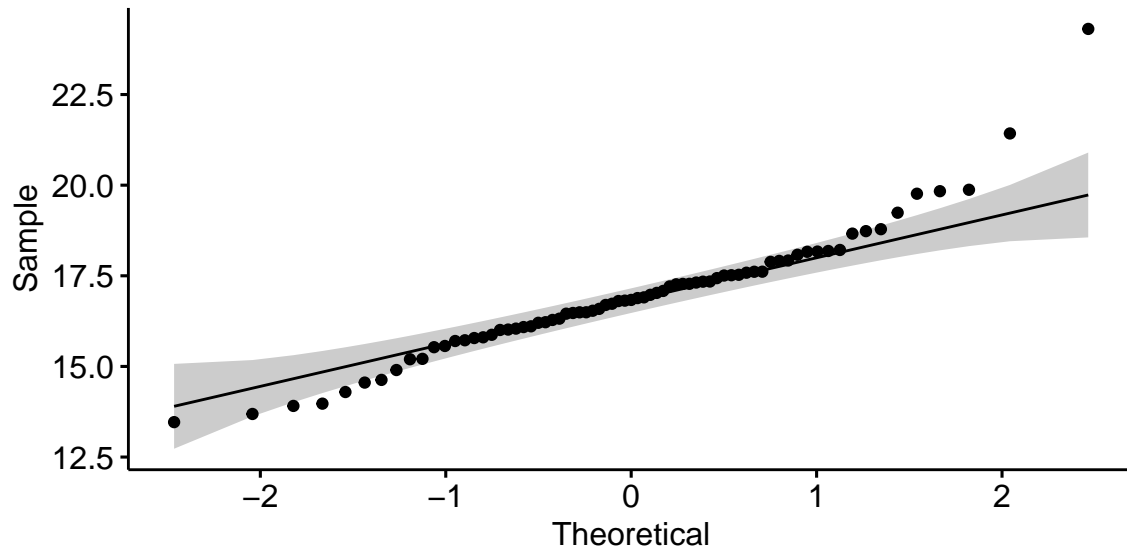


Figure 3. QQ plot of mean tree height data.

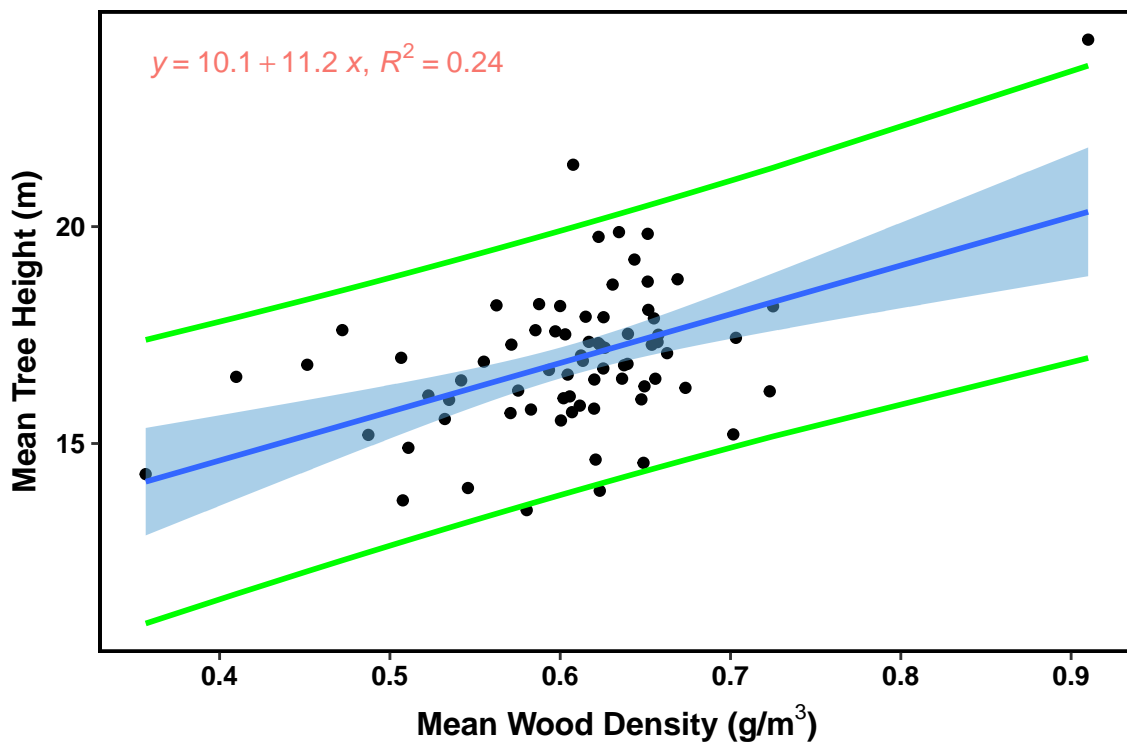


Figure 4. Linear regression of mean wood density and mean tree height with confidence and prediction intervals.

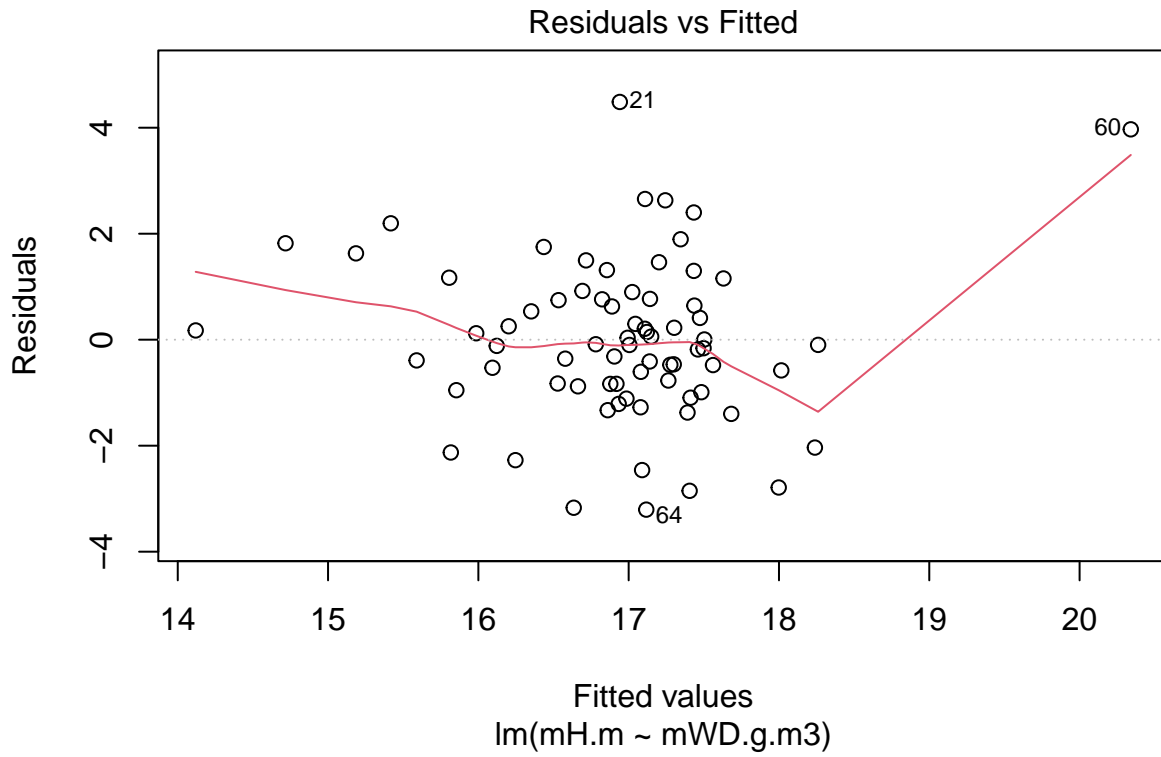


Figure 5. Residuals vs. Fitted values plot of the linear model between mean wood density and mean tree height.

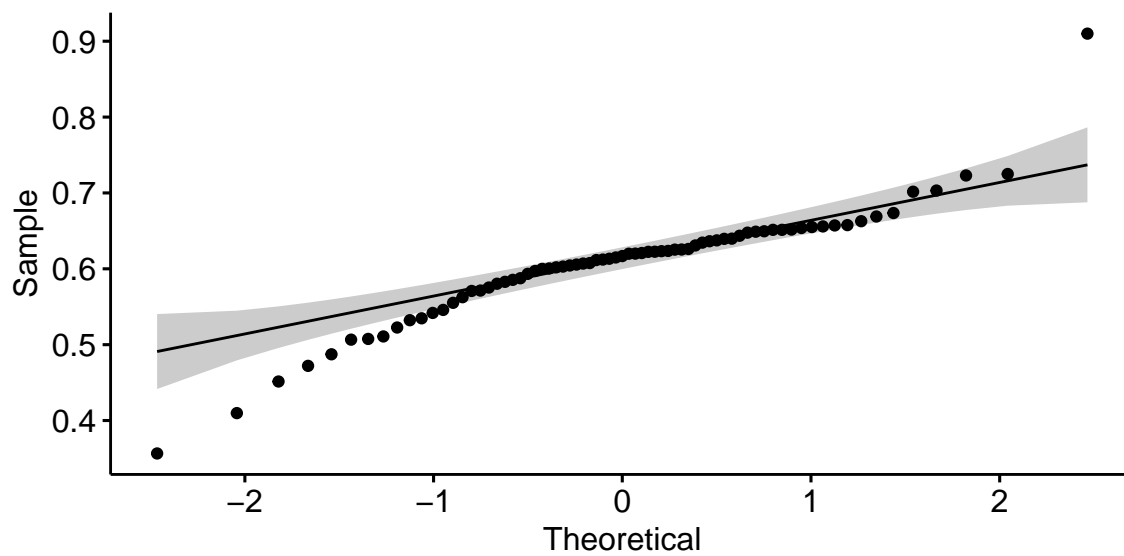


Figure 6. QQ plot of mean wood density data.

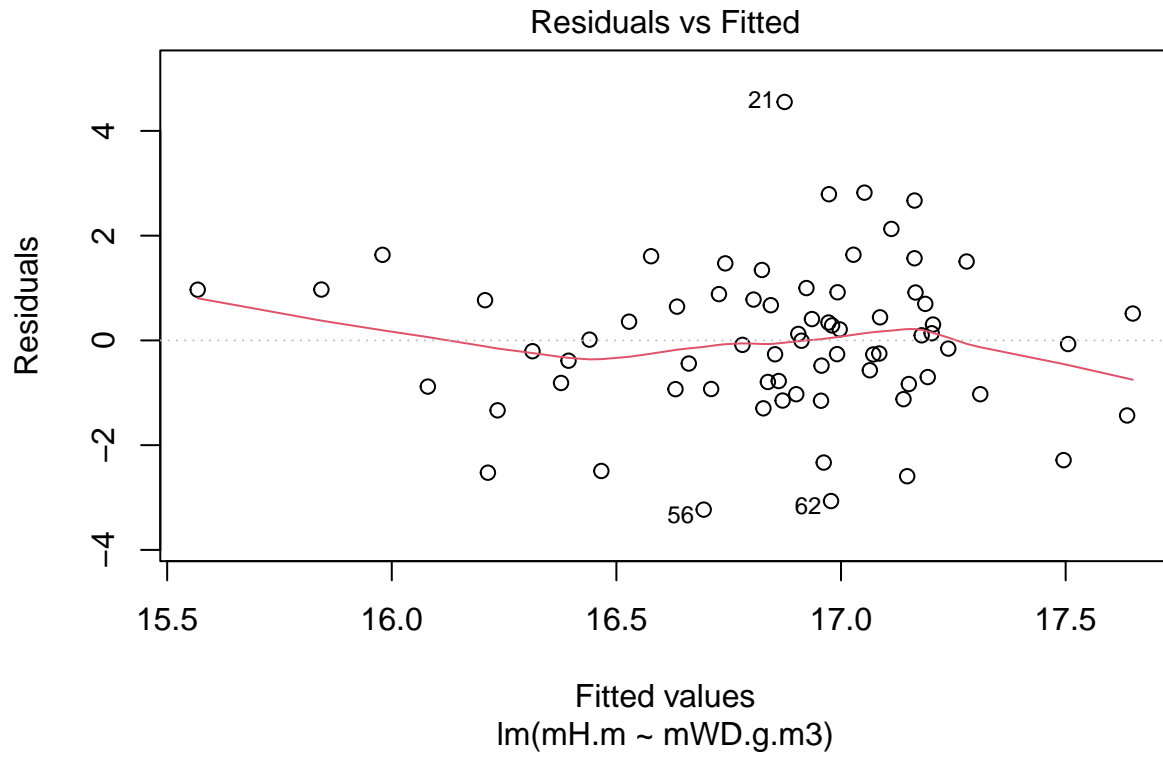


Figure 7. Residuals vs. Fitted values plot of the linear model between mean wood density data after removing outliers.