NDVI Linear regression model

#import the data  
library(readr)  
ndvi <- read\_csv("data/ndvi.csv")

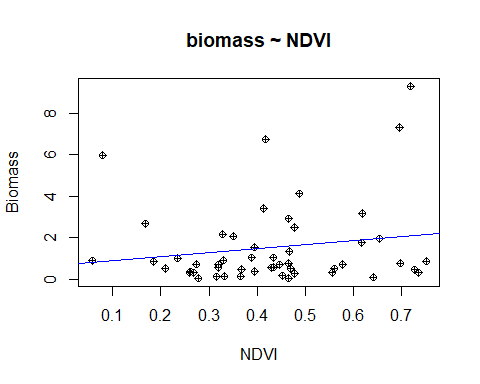
##   
## -- Column specification --------------------------------------------------------  
## cols(  
## num = col\_double(),  
## x = col\_double(),  
## y = col\_double(),  
## NDVI = col\_double(),  
## Biomass = col\_double()  
## )

head(ndvi) #to view the head of th data

## # A tibble: 6 x 5  
## num x y NDVI Biomass  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 248841 1631803 0.413 3.40   
## 2 2 248559 1632085 0.560 0.501  
## 3 3 248841 1632085 0.696 7.31   
## 4 4 249123 1632085 0.452 0.155  
## 5 5 249405 1632085 0.618 3.16   
## 6 6 249687 1632085 0.167 2.66

we want a linear regression model based on this formula Biomass = B0 + B1\*NDVI

#we can visualise the linear relationship between the NDVI and the BIOMASS with a scatterplot  
plot(x = ndvi$NDVI, y = ndvi$Biomass, main = "biomass ~ NDVI",  
 xlab = "NDVI", ylab = "Biomass",  
 pch = 10, frame = TRUE)  
abline(lm(Biomass ~ NDVI, data = ndvi), col = "blue")



# scatterplot

We can compute the correlation between the two variables

cor.test(ndvi$Biomass, ndvi$NDVI)

##   
## Pearson's product-moment correlation  
##   
## data: ndvi$Biomass and ndvi$NDVI  
## t = 1.1793, df = 49, p-value = 0.244  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1147051 0.4223762  
## sample estimates:  
## cor   
## 0.1661301

Our correlation is too closer to 0. So there is no correlation between the two variables.

[Here](https://statisticsbyjim.com/basics/correlations/#:~:text=A%20correlation%20between%20variables%20indicates,change%20in%20a%20specific%20direction.&text=For%20example%2C%20height%20and%20weight,weight%20also%20tends%20to%20increase.) we find how to explain the correlations

We can force to make the linear model but it will not be really significant. But anyway let’s try.

#Our formula is Biomass = B0 + B1\*NDVI  
mbao\_model <- lm(Biomass ~ NDVI , data = ndvi)# the formula  
#we will print the model  
mbao\_model

##   
## Call:  
## lm(formula = Biomass ~ NDVI, data = ndvi)  
##   
## Coefficients:  
## (Intercept) NDVI   
## 0.6774 1.9443

Now that we have built the linear model, we also have established the relationship between the predictor and response in the form of a mathematical formula for biomass as a function for ndvi. For the above output, you can notice the ‘Coefficients’ part having two components: *Intercept*: 0.6774, *NDVI*: 1.9443. These are also called the beta coefficients. In other words,  
Biomass= Intercept + (β ∗ NDVI)

Our model is:

Biomass = 0.6674 + 1.9443\* NDVI

Before using our regression model, we have to ensure that it is statistically significant. We will print it summary to check.

summary(mbao\_model)

##   
## Call:  
## lm(formula = Biomass ~ NDVI, data = ndvi)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8315 -1.1687 -0.7502 0.1010 7.2154   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.6774 0.7583 0.893 0.376  
## NDVI 1.9443 1.6487 1.179 0.244  
##   
## Residual standard error: 1.98 on 49 degrees of freedom  
## Multiple R-squared: 0.0276, Adjusted R-squared: 0.007754   
## F-statistic: 1.391 on 1 and 49 DF, p-value: 0.244

We only have to check the p-value for the statistical significance. The p-Value is very important because, We can consider a linear model to be statistically significant only when both these p-Values are less that the pre-determined statistical significance level, which is ideally 0.05.

In our case the p-value is 0.244, greater than 0.05.

Our model is not significant statistically.