

# M12 Group Work: Simple Linear Regression for TAs

Pablo E. Gutiérrez-Fonseca

2024-04-18

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tibble' was built under R version 4.3.1
```

```
## Warning: package 'tidyr' was built under R version 4.3.2
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
## Warning: package 'purrr' was built under R version 4.3.1
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
## Warning: package 'forcats' was built under R version 4.3.2
```

```
## Warning: package 'lubridate' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.4
```

```
## v forcats   1.0.0      v stringr   1.5.1
```

```
## v ggplot2   3.5.0      v tibble    3.2.1
```

```
## v lubridate 1.9.3      v tidyr     1.3.1
```

```
## v purrr     1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.1
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.1
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.1
```

```
##
```

```
## Attaching package: 'car'
```

```
##
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      recode
```

```
##
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

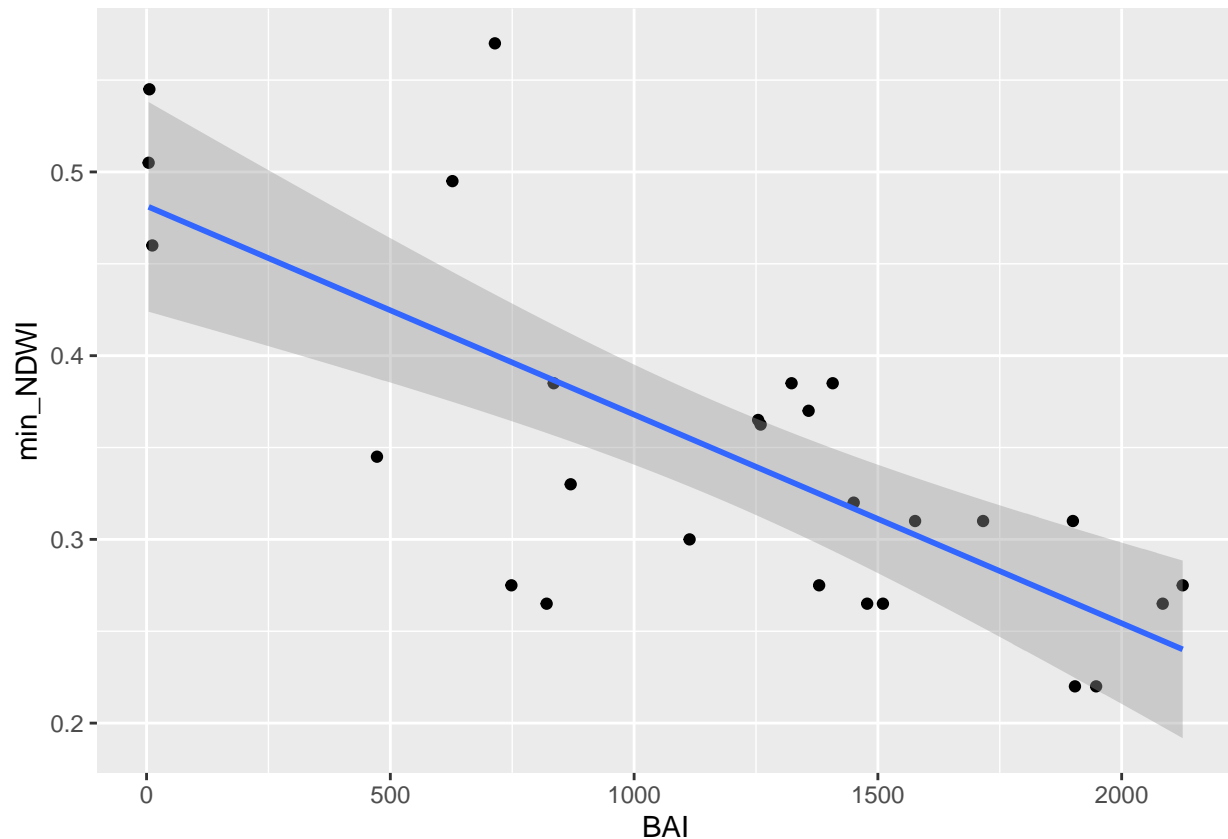
```
##      some
```

```
#Load data
```

```
BAI_NDWI <- read_xlsx("PIRU_BAI.xlsx")
```

```
#This makes a plot with our regression line.
```

```
ggplot(BAI_NDWI) +  
  aes(x = BAI, y = min_NDWI) +  
  geom_point() +  
  theme(legend.position = "none") +  
  stat_smooth(method = "lm", formula = y ~ x, geom = "smooth")
```



#this simply creates a 'model' for our regression model #This is the main think we are looking at!

```
model <- lm(BAI ~ min_NDWI, data = BAI_NDWI)
summary(model)
```

```
##
## Call:
## lm(formula = BAI ~ min_NDWI, data = BAI_NDWI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -777.4 -263.6  138.9  302.8  599.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2835.1      315.7   8.981 2.68e-09 ***
## min_NDWI      -4761.5      877.7  -5.425 1.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 426.4 on 25 degrees of freedom
## Multiple R-squared:  0.5407, Adjusted R-squared:  0.5223
## F-statistic: 29.43 on 1 and 25 DF, p-value: 1.246e-05
```

##the Breusch-Pagan test to assess homoscedasticity

```
ncvTest(model)
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 0.005762692, Df = 1, p = 0.93949
```

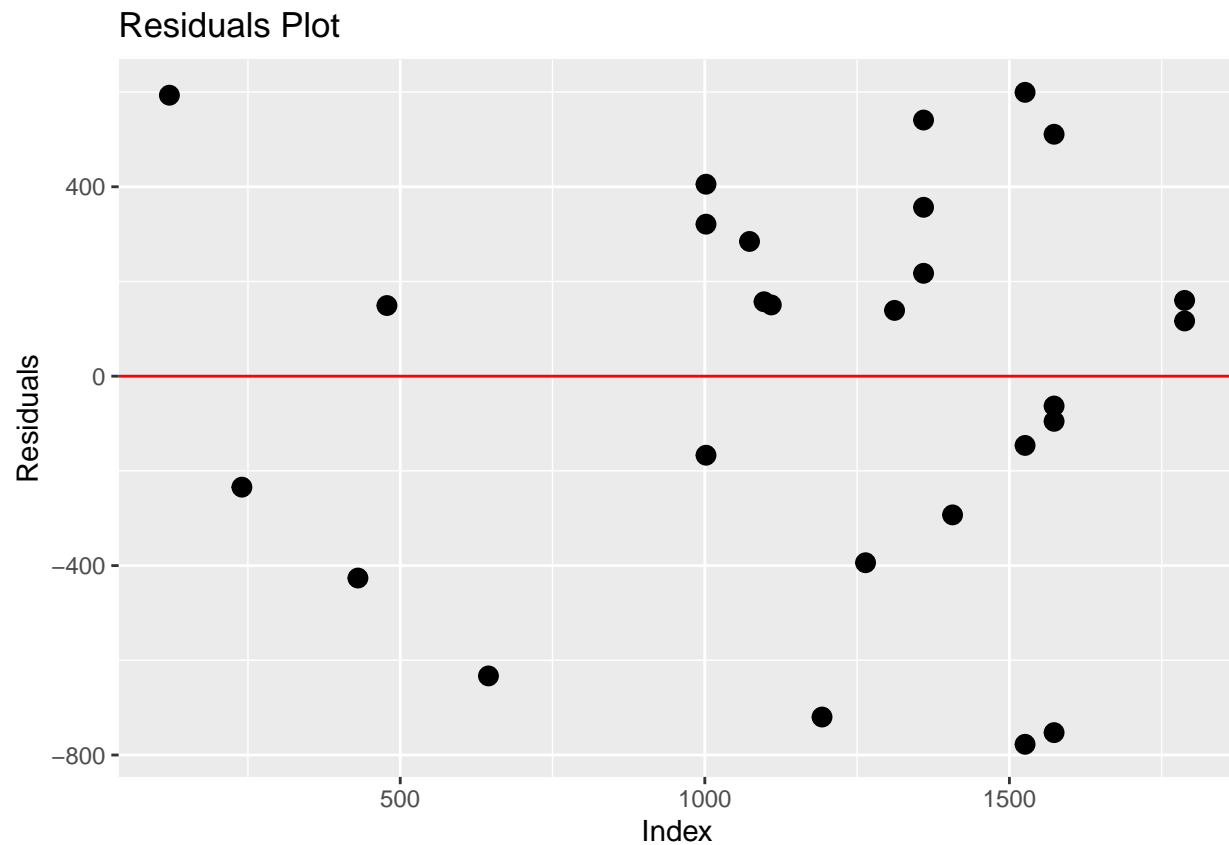
```
#get list of residuals to nest normality
```

```
res <- resid(model)  
shapiro.test(res)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  res  
## W = 0.94094, p-value = 0.1285
```

```
#create residual plot
```

```
ggplot(model, aes(x = .fitted, y = .resid)) +  
  geom_point(size=3) +  
  geom_hline(yintercept = 0, color = "red") +  
  labs(title = "Residuals Plot",  
       x = "Index",  
       y = "Residuals")
```



```
#RMSE
```

```
RMSE <- sqrt(sum(model$residuals^2)/model$df)
```

```
response <- BAI_NDWI$BAI
# Calculate the Mean of the response
mean_response <- mean(response)
Percentage_Error <- RMSE/mean_response
Percentage_Error
```

```
## [1] 0.3609371
```

```
#PRESS
```

```
# (r <- resid(mod))
pr <- resid(model)/(1 - lm.influence(model)$hat)
PRESS <- sum(pr^2)
PRESS
```

```
## [1] 5406891
```

```
# PRESS_RMSE
PRESS_RMSE <- sqrt(PRESS/length(model$residuals))
PRESS_RMSE
```

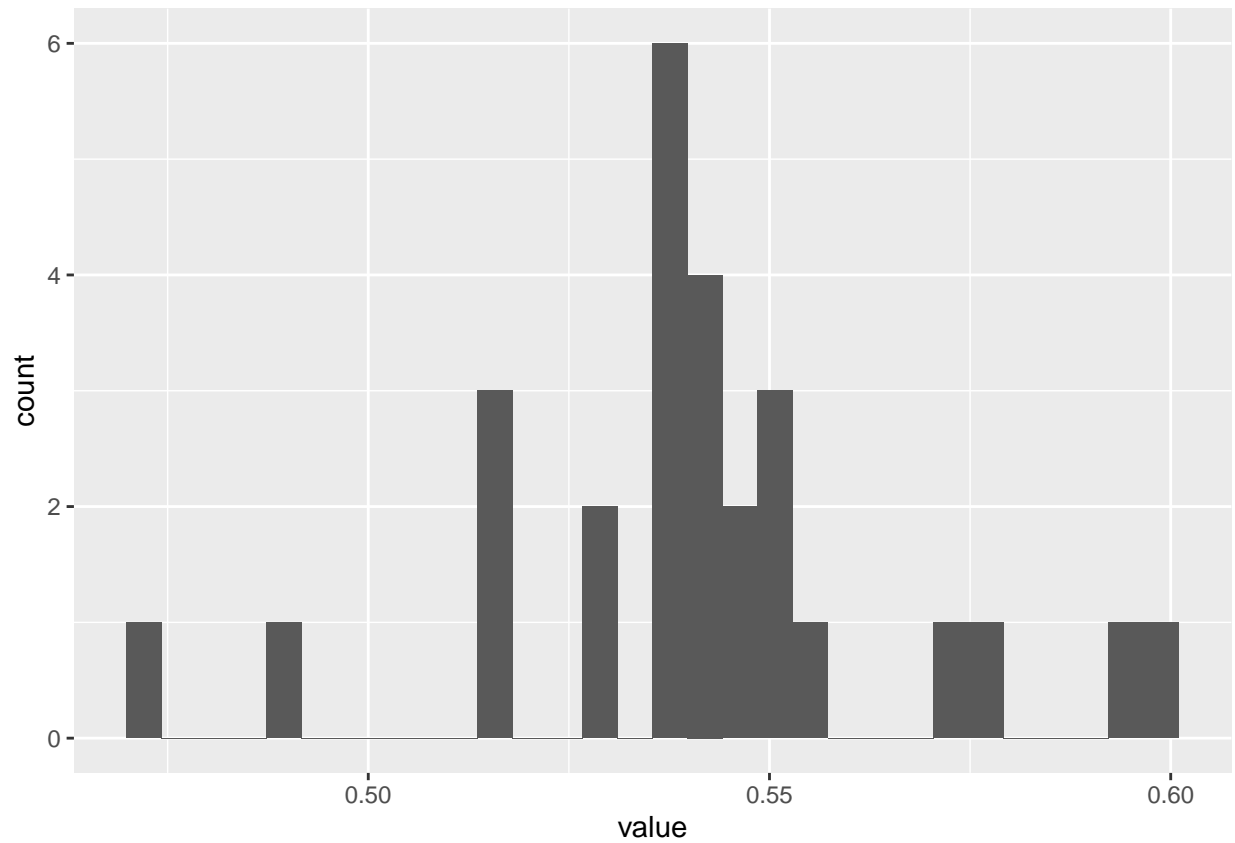
```
## [1] 447.4988
```

```
#Jackknifing
jack.reg<-numeric(27)
for (i in 1:27) {
  model<-lm(BAI[-i]~min_NDWI[-i], data = BAI_NDWI)
  jack.reg[i]<-summary(model)$r.squared }

ggplot(as.tibble(jack.reg), aes(value)) + geom_histogram()
```

```
## Warning: 'as.tibble()' was deprecated in tibble 2.0.0.
## i Please use 'as_tibble()' instead.
## i The signature and semantics have changed, see '?as_tibble'.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
shapiro.test(jack.reg)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  jack.reg  
## W = 0.93286, p-value = 0.08125
```