analysis for erosion

Yingmai Chen,Yan Wang

2023-11-27

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.1     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## -- 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
```

```
erosion<-read.csv("erosion.csv")
```

linear regression for all variables

```
library(readr)
data <- read_csv("erosion.csv")
```

```
## Rows: 31 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Bluff
## dbl (8): Orientation (deg), RR (m/yr), Wave Height for NNE wind 15 m/s (m), ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
names(data) <- gsub(" ", "_", names(data))
names(data) <- gsub("\\(", "", names(data))
names(data) <- gsub("\\)", "", names(data))
names(data) <- gsub("/", "_per_", names(data))
names(data) <- gsub("%", "percent", names(data))
if ("RR (m/yr)" %in% names(data)) {
  names(data)[names(data) == "RR (m/yr)"] <- "RR_m_per_yr"
}
predictors <- setdiff(names(data), c("Bluff", "RR_m_per_yr"))
```

```r
target <- "RR_m_per_yr"
model_formula <- as.formula(paste(target, "~", paste(predictors, collapse = " + ")))
model <- lm(model_formula, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = model_formula, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41517 -0.09785 -0.03958  0.05001  0.48355
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.2239981  0.2468061   0.908  0.37351
## Orientation_deg                     0.0001316  0.0004310   0.305  0.76289
## Wave_Height_for_NNE_wind_15_m_per_s_m  1.3705939  0.4862480   2.819  0.00974 **
## Max_Wave_Height_m                  -0.6028517  0.4346898  -1.387  0.17878
## Mud_percent                        -0.0026008  0.0198475  -0.131  0.89688
## BaseEl_m                           -0.0307893  0.0712946  -0.432  0.66986
## BluffEl_m                          -0.0277326  0.0086660  -3.200  0.00398 **
## Seawall                            -0.0714937  0.1389836  -0.514  0.61188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2326 on 23 degrees of freedom
## Multiple R-squared:  0.612,  Adjusted R-squared:  0.4939
## F-statistic: 5.183 on 7 and 23 DF,  p-value: 0.001192
```

```r
predictions <- predict(model, data)
mse <- mean((data[[target]] - predictions)^2)
rsquared <- summary(model)$r.squared
cat("Mean Squared Error (MSE):", mse, "\n")
```
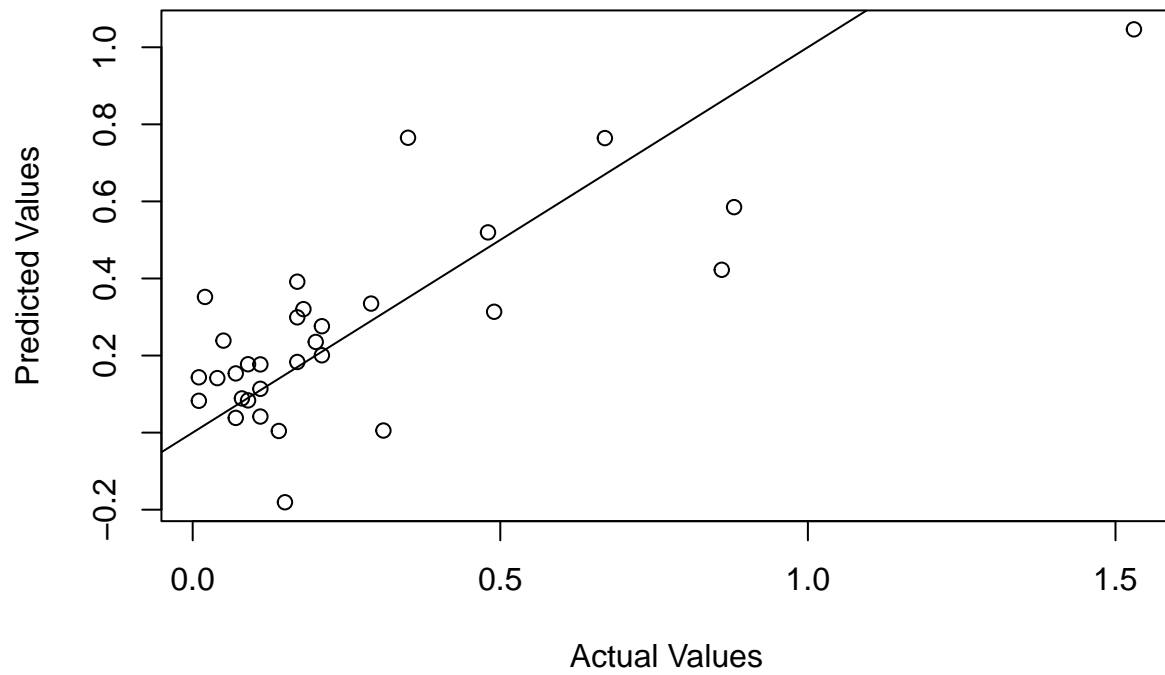
```
## Mean Squared Error (MSE): 0.04013701
```

```r
cat("R-squared:", rsquared, "\n")
```
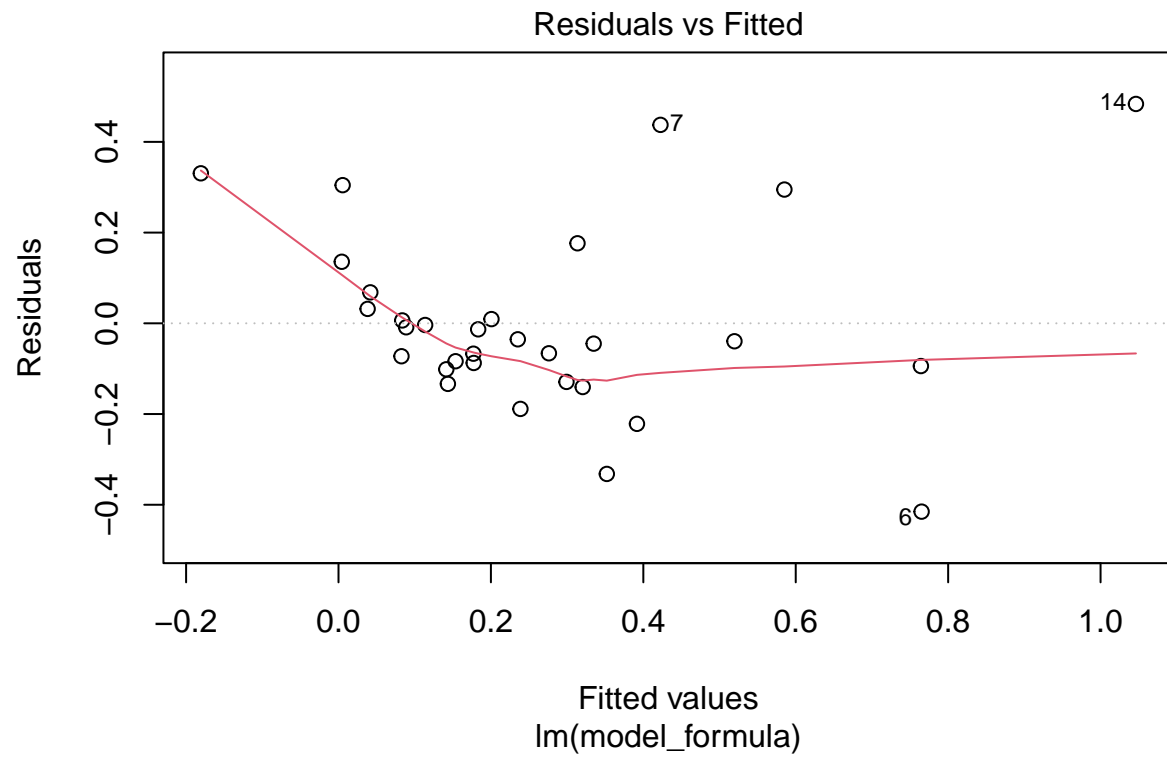
```
## R-squared: 0.6119996
```

check

```r
#Check Linearity
plot(data[[target]], predictions, main = "Fitted vs. Actuals",
     xlab = "Actual Values", ylab = "Predicted Values")
abline(0, 1)
```
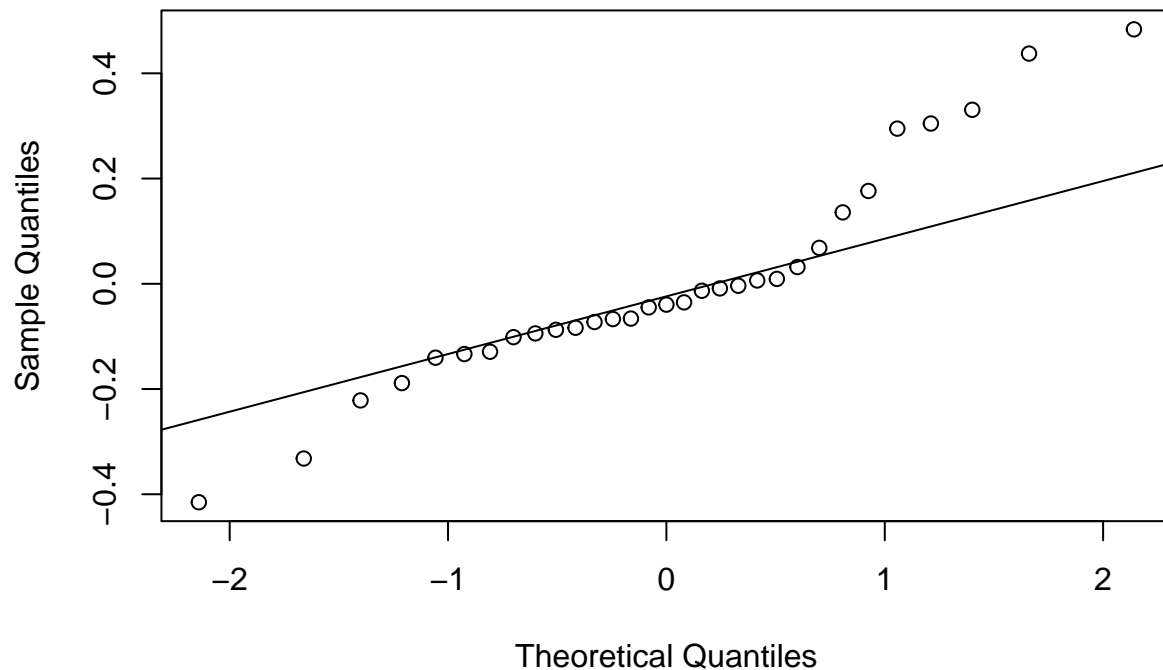
Fitted vs. Actuals


```r
#Check Homoscedasticity
plot(model, which = 1)
```

Residuals vs Fitted

Residuals

Fitted values
lm(model_formula)

```r
#Check Normal Distribution of Residuals
qqnorm(model$residuals)
qqline(model$residuals)
```

Normal Q–Q Plot

factor analysis for nominal variable

```r
data <- read.csv("erosion.csv")
data$Seawall <- factor(data$Seawall)
model <- lm(`RR..m.yr.` ~ Seawall, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = RR..m.yr. ~ Seawall, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44100 -0.13643 -0.07143  0.02879  1.07900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.18143    0.06676   2.717   0.0110 *
## Seawall1     0.26957    0.11755   2.293   0.0293 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3059 on 29 degrees of freedom
## Multiple R-squared:  0.1535, Adjusted R-squared:  0.1243
```

```
## F-statistic: 5.259 on 1 and 29 DF,  p-value: 0.02927
```

```
ggplot(data, aes(x = Seawall, y = `RR..m.yr.`)) +
  geom_boxplot() +
  labs(title = "Effect of Seawall on RR (m/yr)", x = "Seawall", y = "RR (m/yr)")
```

Effect of Seawall on RR (m/yr)