Week 7 Workshop

Katie Miller

2023-05-17

library(tidyverse)  
library(here)  
library(lterdatasampler)  
  
library(performance)  
library(broom)  
library(flextable)  
library(ggeffects)  
library(car)  
library(naniar)

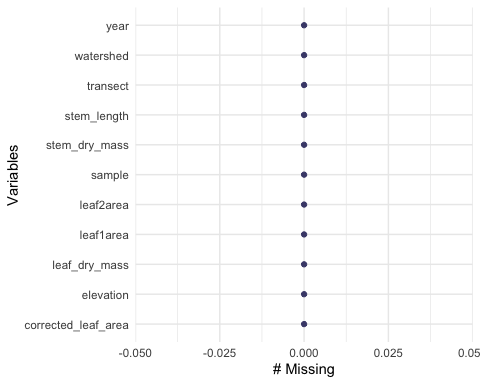
# Linear Models

How does stem length *predict* stem dry mass?

maples\_data <- hbr\_maples %>%   
 filter(year == 2003 & watershed == "Reference")

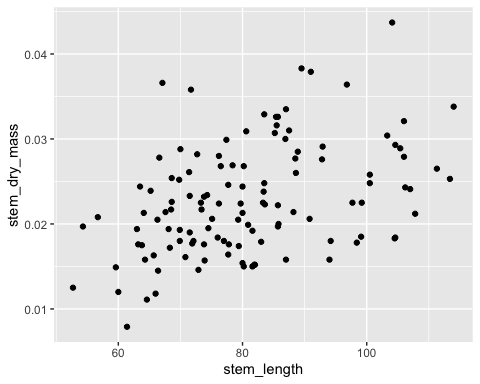
Visualize missing data

gg\_miss\_var(maples\_data)



Exploratory data visualization

ggplot(data = maples\_data, aes(x= stem\_length, y = stem\_dry\_mass)) +  
 geom\_point()



Let’s try a model

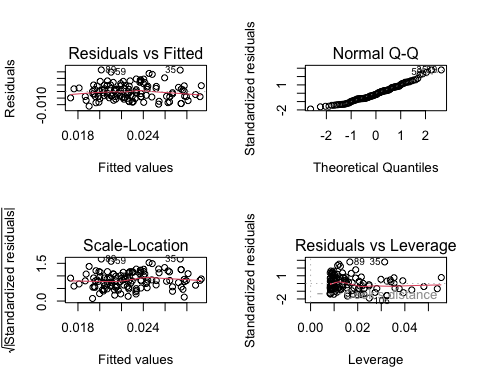
maples\_model <- lm(stem\_dry\_mass ~ stem\_length,  
 data = maples\_data)  
  
maples\_model

##   
## Call:  
## lm(formula = stem\_dry\_mass ~ stem\_length, data = maples\_data)  
##   
## Coefficients:  
## (Intercept) stem\_length   
## 0.0070033 0.0001958

Check assumptions

I. linear relationship: yes! Plot in console II. independence of errors: yes! (based on data collection) III. Homoskedasicity of errors: yes! (residuals vs. fitted plot) IV. Normally distrobuted errors: yes! (QQ plot of residuals)

par(mfrow = c(2, 2))  
plot(maples\_model)



turn off 2 x 2 grid

dev.off()

## null device   
## 1

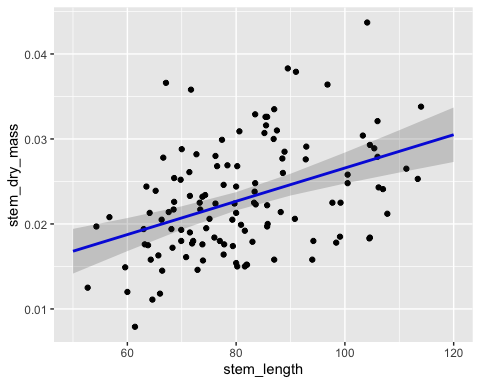
# Putting together to communicate

## model predictions

prediction <- ggpredict(maples\_model, terms = "stem\_length")

plot predictions

plot\_prediction <- ggplot(data = maples\_data, aes(x = stem\_length, y = stem\_dry\_mass)) +  
 geom\_point() +  
 geom\_line(data = prediction, aes(x = x, y = predicted), color = "blue", linewidth = 1) +  
 #confidence interval  
 geom\_ribbon(data = prediction, aes(x=x, y = predicted, ymin = conf.low, ymax = conf.high), alpha = 0.2)  
  
plot\_prediction



# same as previous plot

# Create tables

model\_summary <- summary(maples\_model)  
  
model\_squares <- anova(maples\_model)

making a table

model\_squares\_table <- tidy(model\_squares) %>%   
 # round the sum of squares and mean squares columns to have 5 digits (could be less)  
 mutate(across(sumsq:meansq, ~ round(.x, digits = 5))) %>%   
 # round the F-statistic to have 1 digit  
 mutate(statistic = round(statistic, digits = 1)) %>%   
 # replace the very very very small p value with < 0.001  
 mutate(p.value = case\_when(  
 p.value < 0.001 ~ "< 0.001"  
 )) %>%   
 # rename the stem\_length cell to be meaningful  
 mutate(term = case\_when(  
 term == "stem\_length" ~ "Stem length (mm)",  
 TRUE ~ term  
 )) %>%   
 # make the data frame a flextable object  
 flextable() %>%   
 # change the header labels to be meaningful  
 set\_header\_labels(df = "Degrees of Freedom",   
 sumsq = "Sum of squares",  
 meansq = "Mean squares",  
 statistic = "F-statistic",  
 p.value = "p-value")  
  
model\_squares\_table

| term | Degrees of Freedom | Sum of squares | Mean squares | F-statistic | p-value |
| --- | --- | --- | --- | --- | --- |
| Stem length (mm) | 1 | 0.00089 | 0.00089 | 25.1 | < 0.001 |
| Residuals | 118 | 0.00417 | 0.00004 |  |  |