Homework 03

Sarah Tate

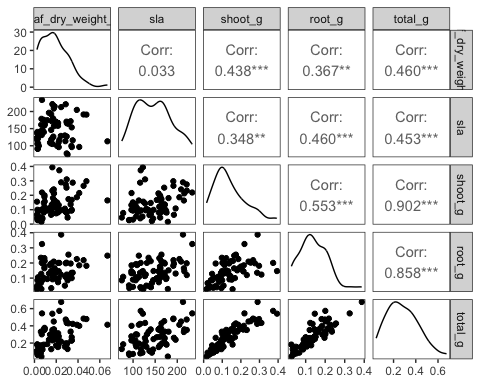
Table of contents

# general use  
library(tidyverse)  
library(readxl)  
library(here)  
library(janitor)  
  
# visualizing pairs  
library(GGally)  
  
# model selection  
library(MuMIn)  
  
# model predictions  
library(ggeffects)  
  
# model tables  
library(gtsummary)  
library(flextable)  
library(modelsummary)  
  
drought\_exp <- read\_xlsx(path = here("data",   
 "Valliere\_etal\_EcoApps\_Data.xlsx"),  
 sheet = "First Harvest")

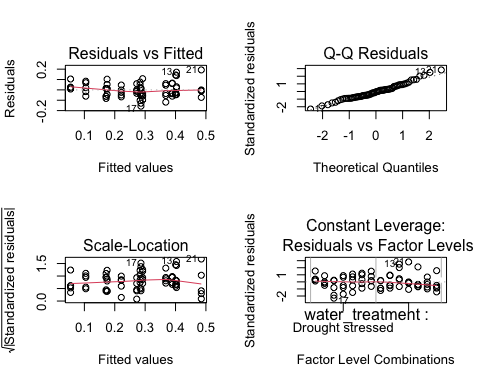
# cleaning  
drought\_exp\_clean <- drought\_exp %>%   
 clean\_names() %>% # nicer column names  
 mutate(species\_name = case\_when( # adding column with species scientific names  
 species == "ENCCAL" ~ "Encelia californica", # bush sunflower  
 species == "ESCCAL" ~ "Eschscholzia californica", # California poppy  
 species == "PENCEN" ~ "Penstemon centranthifolius", # Scarlet bugler  
 species == "GRICAM" ~ "Grindelia camporum", # great valley gumweed  
 species == "SALLEU" ~ "Salvia leucophylla", # Purple sage  
 species == "STIPUL" ~ "Nasella pulchra", # Purple needlegrass  
 species == "LOTSCO" ~ "Acmispon glaber" # deerweed  
 )) %>%   
 relocate(species\_name, .after = species) %>% # moving species\_name column after species  
 mutate(water\_treatment = case\_when( # adding column with full treatment names  
 water == "WW" ~ "Well watered",  
 water == "DS" ~ "Drought stressed"  
 )) %>%   
 relocate(water\_treatment, .after = water) # moving water\_treatment column after water

#correlations

ggpairs(drought\_exp\_clean, # data frame  
 columns = c("leaf\_dry\_weight\_g", # columns to visualize  
 "sla",   
 "shoot\_g",   
 "root\_g",   
 "total\_g"),   
 upper = list(method = "pearson")) + # calculating Pearson correlation coefficient  
 theme\_bw() + # cleaner theme  
 theme(panel.grid = element\_blank()) # getting rid of gridlines



# null model  
model0 <- lm(total\_g ~ 1, # formula  
 data = drought\_exp\_clean) # data frame  
  
# saturated model  
model1 <- lm(total\_g ~ sla + water\_treatment + species\_name,  
 data = drought\_exp\_clean)  
  
# model 2: total mass as predicted by specific leaf area and water treatment  
model2 <- lm(total\_g ~ sla + water\_treatment,  
 data = drought\_exp\_clean)  
  
# model 3: total mass as predicted by specific leaf area and species name  
model3 <- lm(total\_g ~ sla + species\_name,  
 data = drought\_exp\_clean)  
  
# model 4: total mass as predicted by species name and water treatments  
model4 <- lm(total\_g ~ water\_treatment + species\_name,  
 data = drought\_exp\_clean)  
  
#diagnostics for model 4  
par(mfrow = c(2, 2))  
plot(model4)



model.sel(model0,  
 model1,   
 model2,   
 model3,  
 model4)

Model selection table   
 (Int) sla spc\_nam wtr\_trt df logLik AICc delta weight  
model4 0.05455 + + 9 88.598 -156.2 0.00 0.772  
model1 0.07994 -0.0002475 + + 10 88.741 -153.8 2.44 0.228  
model3 -0.03315 0.0012900 + 9 72.538 -124.1 32.12 0.000  
model2 0.04670 0.0012810 + 4 52.220 -95.8 60.37 0.000  
model0 0.27900 2 39.580 -75.0 81.22 0.000  
Models ranked by AICc(x)

MODEL 4 IS BEST

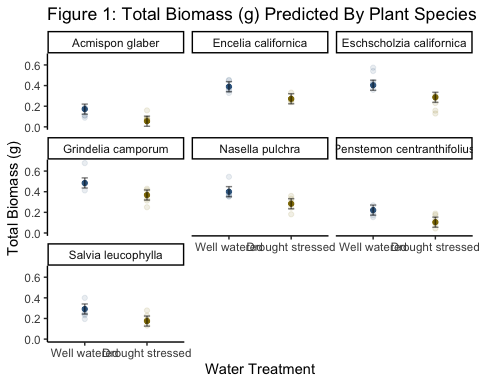
summary(model4)

Call:  
lm(formula = total\_g ~ water\_treatment + species\_name, data = drought\_exp\_clean)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.157087 -0.046953 -0.003733 0.041244 0.192657   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.05455 0.02451 2.225 0.02973 \*   
water\_treatmentWell watered 0.11695 0.01733 6.746 5.90e-09 \*\*\*  
species\_nameEncelia californica 0.21774 0.03243 6.714 6.70e-09 \*\*\*  
species\_nameEschscholzia californica 0.23164 0.03243 7.143 1.22e-09 \*\*\*  
species\_nameGrindelia camporum 0.31335 0.03243 9.662 5.53e-14 \*\*\*  
species\_nameNasella pulchra 0.22881 0.03243 7.055 1.72e-09 \*\*\*  
species\_namePenstemon centranthifolius 0.05003 0.03243 1.543 0.12799   
species\_nameSalvia leucophylla 0.12020 0.03243 3.706 0.00045 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.07252 on 62 degrees of freedom  
Multiple R-squared: 0.7535, Adjusted R-squared: 0.7257   
F-statistic: 27.08 on 7 and 62 DF, p-value: < 2.2e-16

Problem 1: Multiple linear regression - model selection and construction

1. MODEL 4 IS BEST BASED ON TEST, CHECK DIAGNOSTICS

model\_preds <- ggpredict(model4,   
 terms = c("water\_treatment",   
 "species\_name"))  
  
model\_preds\_for\_plotting <- model\_preds %>%   
 rename(water\_treatment = x, # renaming columns to make this easier to use  
 species\_name = group)  
  
  
ggplot() +  
 # underlying data  
 geom\_point(data = drought\_exp\_clean,  
 alpha = 0.1,  
 aes(x = water\_treatment,  
 y = total\_g,  
 color = water\_treatment)) +  
 # model prediction 95% CI ribbon  
 geom\_point(data = model\_preds\_for\_plotting,  
 aes(x = water\_treatment,   
 y = predicted,  
 color = water\_treatment),  
 alpha = 1) +  
 # model prediction lines  
 geom\_errorbar(data = model\_preds\_for\_plotting,  
 aes(x = water\_treatment,   
 y = predicted,  
 ymin = conf.low,   
 ymax = conf.high,  
 fill = water\_treatment),  
 width = 0.1,  
 alpha = 0.5) +  
 # cleaner theme  
 theme\_classic() +  
 # creating different panels for species  
 facet\_wrap(~species\_name)+   
 theme(legend.position = "none")+  
 labs(x = "Water Treatment",  
 y = "Total Biomass (g)",  
 title = "Figure 1: Total Biomass (g) Predicted By Plant Species And Water Treatment") +  
 scale\_color\_manual(values = c("Well watered" = "steelblue4",  
 "Drought stressed" = "gold4"))



Problem 2: Affective visualization

Problem 3: Statistical critique