Homework 3 Submission Document

Jack Phillips

2024-05-28

Table of contents

Link to Forked Repository

Problem 1. Multiple Linear Regression: Model Selection and Construction (52 points)

a. Make a table or list of all the models from class and the last one you constructed on your own. Write a caption for your table. (8 points)

Caption: table captions typically go above the table. Number the table and provide a title. Describe what is in the table (columns and rows).

Table: In your table, each row should be a model with the model number (1, 2, 3, etc.) and the predictors for each model.

Load packages and data

```
# general use
library(tidyverse)
library(readxl)
library(here)
library(janitor)
```

```r
# 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")

# quick look at data
str(drought_exp)
```

```
tibble [70 x 13] (S3: tbl_df/tbl/data.frame)
 $ Species            : chr [1:70] "ENCCAL" "ENCCAL" "ENCCAL" "ENCCAL" ...
 $ Water              : chr [1:70] "WW" "WW" "WW" "WW" ...
 $ Rep #              : num [1:70] 1 2 3 4 5 1 2 3 4 5 ...
 $ Height (cm)        : num [1:70] 5.8 4.9 8.4 6.5 7.1 3.2 4.4 4.2 4.5 3.9 ...
 $ Leaf #             : num [1:70] 11 8 11 12 10 7 7 10 8 6 ...
 $ Leaf dry weight (g): num [1:70] 0.0294 0.0185 0.0177 0.0178 0.0164 0.017 0.0193 0.0153 0.0
 $ Leaf area (cm2)    : num [1:70] 5.01 3.98 3.69 3.84 3.63 3.06 3.1 2.94 2.73 2.61 ...
 $ SLA                : num [1:70] 170 215 209 216 222 ...
 $ Total LA           : num [1:70] 55.1 31.8 40.6 46.1 36.3 ...
 $ Shoot (g)          : num [1:70] 0.253 0.164 0.241 0.213 0.232 ...
 $ Root (g)           : num [1:70] 0.202 0.165 0.209 0.146 0.12 ...
 $ Total (g)          : num [1:70] 0.455 0.329 0.45 0.359 0.352 ...
 $ R:S                : num [1:70] 0.8 1 0.9 0.7 0.5 0.8 1.2 3.1 0.9 1.2 ...
```

```r
class(drought_exp)
```

```
[1] "tbl_df"     "tbl"        "data.frame"
```

Clean Data

```
# 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

# Borrowed from template
```

Define & Compare Models

```
model0 <- lm(total_g ~ 1, data = drought_exp_clean)
model1 <- lm(total_g ~ sla + water_treatment + species_name, data = drought_exp_clean)
model2 <- lm(total_g ~ sla + water_treatment, data = drought_exp_clean)
model3 <- lm(total_g ~ sla + species_name, data = drought_exp_clean)
model4 <- lm(total_g ~ water_treatment + species_name, data = drought_exp_clean)

model_comparison <- model.sel(model0, model1, model2, model3, model4)
print(model_comparison)
```

```
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)
```

Create a Table Comparing Models

```
models_table <- tibble::tribble(
  ~`Model Number`, ~`Predictors`, ~`AIC`,
  0, "None (Null Model)", AIC(model0),
  1, "SLA, Water Treatment, Species", AIC(model1),
  2, "SLA, Water Treatment", AIC(model2),
  3, "SLA, Species", AIC(model3),
  4, "Water Treatment, Species", AIC(model4)
)
```

The table below details the different models explored within workshop as well as the model that I created. Along with model number and predictors, AIC value for each value is also displayed.

```
print(models_table)
```

```
# A tibble: 5 x 3
  `Model Number` Predictors                      AIC
           <dbl> <chr>                           <dbl>
1              0 None (Null Model)               -75.2
2              1 SLA, Water Treatment, Species -157.
3              2 SLA, Water Treatment            -96.4
4              3 SLA, Species                   -127.
5              4 Water Treatment, Species       -159.
```