

Module 4 Assessment

Include your name

Instructions

Please submit the markdown file and compiled pdf to canvas before Mar 29 at 11:59pm. For this assignment, you can discuss with classmates, but please at least attempt to go through it individually first so that you can see what you understand or don't understand. Ultimately, the final product you turn in should be your own work. So you can discuss questions with classmates, but your answers should be written in your own words.

Intro

In this module assessment, you'll be considering data from "Soil nutrients influence growth response tree species to drought" by Levesque, Walthert, and Weber (Journal of Ecology, 2016). The authors consider how certain properties of soil and climate are associated with tree growth. The data we will be using today can be accessed at this link: <https://datadryad.org/stash/dataset/doi:10.5061%2Fdryad.sd40d>, but can also be taken directly from my github using the code chunk below.

In particular, the authors consider Basal area increment (BAI) (i.e., a measure of tree cross section growth) as the dependent variable and measure 538 trees across 52 sites in Switzerland and northwestern Italy. Measurements were taken on the same trees annually from 1957-2006. Take a few minutes to read skim the study (at least the abstract) to get a sense for the scientific question of interest.

```
## load libraries which we will need later
library("sandwich")
library("lmtest")
library("lme4")
library("lmboot")

# Load in the data
tree_growth <- read.csv("https://raw.githubusercontent.com/ysamwang/btry6020_sp22/main/lectureData/Trees")
# Remove data points with missing data
tree_growth <- na.omit(tree_growth)

## Dimensions of data set
dim(tree_growth)

## [1] 26627    17

# variables in data set
names(tree_growth)

## [1] "year"    "site"    "siteid"  "species" "treeid"  "trw"     "bai"
## [8] "radius"  "age"     "awc100"  "pH"      "BS"      "C.N"     "tmean"
## [15] "prec"    "dri"     "co2"
```

Question 1 (2 pts)

The authors are interested in the dependent variable Basal area increment which is in the column named `bai`. Specifically, they take the log transform (and add 1 so that they avoid taking the log of 0) and ultimately use $\log(\text{bai} + 1)$. Fit a linear model which examines the association between $\log(\text{bai} + 1)$ and the covariates: radius (radius of the tree), age, `prec` (which is precipitation in mm) and `co2` (atmospheric CO_2). Don't transform the covariates. Write a sentence about the interpretation of the estimated coefficient corresponding to `prec`. You can interpret the model as if we aren't adding the additional 1 to `bai` inside the log.

Fit the linear model specified above here

Answer to Question 1: interpretation of estimated coefficient

Question 2 (1 pt)

Using the model above, form a 95% confidence interval for the coefficient corresponding to the coefficient corresponding to `prec`.

Calculate the lower and upper part of the CI here

Question 3 (5 pts)

Are the modeling assumptions we typically make when creating confidence intervals and hypothesis tests satisfied? Consider at least two of the required assumptions. Explain why or why not. If helpful, you may provide plots to support your claim.

Answer for Question 3

Question 4 (1 pt)

Using the model above, form a 95% confidence interval for the coefficient corresponding to the coefficient corresponding to `prec` using the robust standard errors (i.e., sandwich standard errors).

Calculate the lower and upper part of the CI using robust standard errors

Question 5 (1 pt)

Do you think it's better to use the robust standard errors or the model based standard errors for this setting? Explain.

Answer for Question 5

Question 6 (1 pt)

Using the pairs bootstrap with the percentile method, create a 95% confidence interval for the estimated coefficient corresponding to `prec`.

Calculate the lower and upper part of the CI using the bootstrap here

Question 7 (Bonus: 1 pt)

Is the pairs bootstrap appropriate for this setting? Why or why not?

Answer for Question 7

Question 8 (2 pt)

The authors use a random effects model to model the data and include a random effect for the specific tree and the site. Explain why the authors might be motivated to do that.

Answer for Question 8

Question 9 (1 pt)

Fit a mixed effects model which uses `log(bai +1)` as the dependent variable and includes fixed effects for age, radius, prec and co2. Include a random effect for the tree (`treeid`) and site (`siteid`).

```
## Fit the mixed effects model here
```

Question 10 (1 pt)

Fit the same model as above, but this time use a fixed effect for each tree and site.

```
## Fit the fixed effects model here
```

Question 11 (2 pts)

The two procedures give similar estimates, and neither model is clearly better than the other. But, explain some reasons why you might pick one over the other.

Answer for Question 11