Assignment #2: Fit and Interpret a Linear or Mixed-Effects Model

Due: Week 7
Weight: 20%

Goal: Use your dataset to fit and interpret a linear or mixed-effects model. You will identify a relevant ecological hypothesis, build and assess your model, and visualize key results.

Assignment Components (Total: 75 points)

1. **Define a Research Question and Hypothesis (10 points)**

- State a clear ecological question related to your dataset.
    - Droughts associated with climate change negatively impact tree productivity (Bharambe et al., 2023), biodiversity (Weiskopf et al., 2020), and ecosystem services , Malhi et al., 2020). As droughts occur more frequently and with greater severity (Gu et al., 2020), understanding the strategies trees use to tolerate drought becomes more critical for the sustainability of natural ecosystems. Widespread tree mortality events have catastrophic effects on carbon and water cycles (Bolan et al., 2024) and the functional stability of ecosystems (Fry et al., 2014; Roeder et al., 2021; Turnbull et al., 2016; W. Wang et al., 2012) largely due to the role of tree tissues in net ecosystem hydraulic exchange. The whole objective of this study is to look at how trees manage water movement and the functional mechanism using measured functional traits as predictors. The response variables in the analysis is tree water deficit (TWD). TWD is a continuous measurement done every 15 minutes. One of the criteria for defining data collection is the ecosystem boundaries where vegetation is more exposed to drought events.
- Identify your response and predictor variables.
    - Response variable is trends from the dendrometer [Stable, decreasing, increasing but we used the corresponding slope from the Sen's slope function]. This gives an estimate of the tree tissue shrinking or swelling also known as the tree water deficit (TWD) which is viewed as a continuous drought mechanism. And my predictor variables are water potential (WP), DBH, tree height, leaf water content (LWC), leaf dry matter content (LDMC) and Stress level.

- Clearly state the hypothesis and expected relationship(s).
    - H1: Using the TWD slope from the Sen's slope, we can associate what functional trait tells us if a tree is experiencing a decline, incline or stable TWD. We expect traits that play a role in plant water relations to show a mechanism for state of plant WP.
    - H2: we expect species that live in the same microclimate to have same adaptive functional trait. We expect Juniper spp. which is drought tolerant to adapt to more negative WP even though its TWD is higher but Pinyon pine which are drought avoidant to have a less negative WP even if its TWD is higher or lower, likewise Ponderosa pine to have less negative WP as the TWD increases.

Fit a Linear or Mixed-Effects Model (15 points)

- Fit a linear model (lm()) or a mixed-effects model (lme() or lmer()), depending on your study design.
    - I fitted a lm model, gls, and mixed model. In the lm and gls model, I assumed that there wasn't any variance in the data set. And I fit a mixed model with species nested in plot, just species and plot.
- Justify your model structure, including any random effects.
    - H1 & H2: I included the species and plot as a random effect because it showed a variance when I displayed the dataset on a boxplot. A high percentage of variance in my dataset was caused by species and plot differences. I chose the nested mixed model because the AIC was lower (nested model AIC = 576.5769) than all other models.

3. **Perform Model Selection (10 points)**

- Use model selection criteria (e.g., AIC, delta AIC, likelihood ratio tests) to justify your top model
    - H1 and H2: Using the drop1() model selection algorithm, we selected the model containing only LWC as the best performing model based on the better AIC score.
- This can include identifying the best random structure and the best fixed structure.

- We used the nested mixed model because the AIC is lower for H1. And for H2 we used the model with species as a fixed effect and plot as a random effect. I noticed the estimate of variables significant from H1 didn't change much.

Interpret the Model Results (10 points)

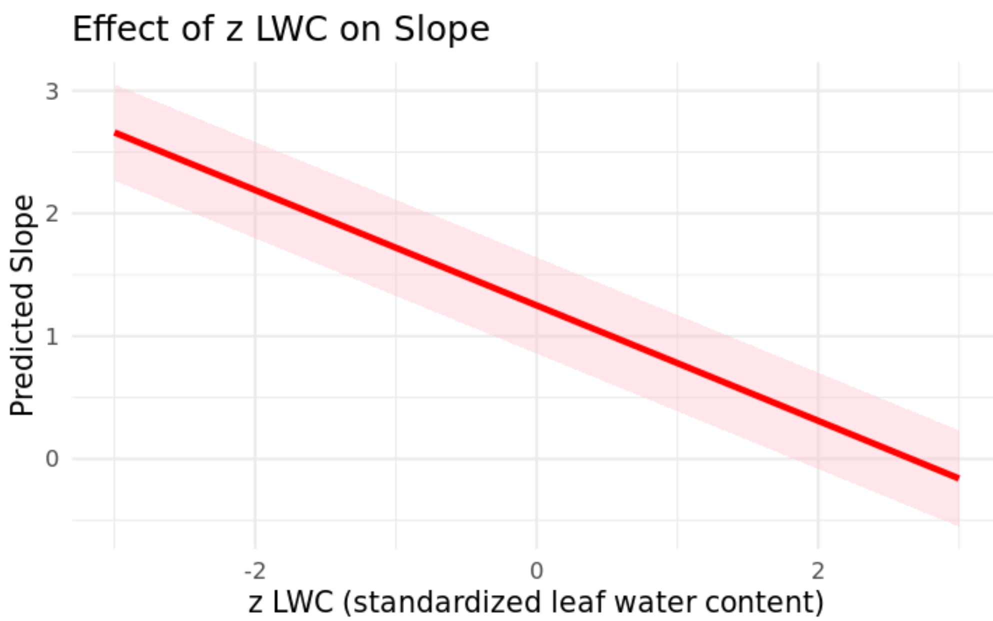
- Interpret coefficients, p-values, and direction/magnitude of effects.
  - H1: Using the best model from model selection, the only significant relationship between our predictors and response variables is LWC. The p value is less than 0.05 (P value = 0.042) which means our model with LWC has a slope different from 0 and it's better than our null model (which is the model without any of the predictors and slope is == 0). The model intercept is -0.89, so when our predictor is at 0, we expect a -0.89 change in the slope of the TWD. The slope of LWC is at -0.47, therefore, for any 1 unit increase in LWC, we expect a decrease by 0.47 in our TWD slope. The interclass correlation coefficient (ICC) is 0.46. This tells us that 46% of the unexplained variance in our model is explained by our random variables which I think might be a bit low.
  - H2: the stress level linear relationship with the TWD slope isn't significant (P value = 0.792), using the visually condition of a tree bark, canopy and stem won't be enough to detect the tree water deficit. Though the linear relationship isn't significant, the cubic and 4th degree are significant (P value = 0.018 & 0.029 respectively). Though they are significant, they have a different effect on the TWD. Stress Level index which goes from 0 (dead) to 10 (healthy). Stress levels 3 and 4 are really evident stressed-out tree. The positive cubic coefficient (2.64) means the middle range of stress values will increase the TWD faster than at the extremes. When the tress level is at 4th degrees, the TWD slope drops at very low (dead) and very high (healthy) stress levels.
  - Lastly, we looked at effect of species on the magnitude of TWD. Out of 12 species, only 4 species had a significant relationship with the slope of TWD. JUOS, PIPO, PIST and lastly QUEM with a P value of 0.001, <0.001, 0.01 and <0.001 respectively. They all have a negative intercept of -4.19, -2.35, -3.58 and -.6.85 for JUOS, PIPO, PIST, QUEM respectively. Therefore, these species reduce the slope of TWD. PIST which is *Pinus strobiformis*, PIPO which is *Pinus ponderosa*, JUOS which is *Juniperus osteosperma* and QUEM which is *Quercus emoryi*. This suggests more stable or improving water status over time.

- Explain the ecological relevance of your results.
  - Overall, our study indicates that as LWC increases, the TWD slope decreases, suggesting that trees with higher leaf water content experience smaller changes in TWD.
  - H1: Ecologically, we found that as the LWC increases, the TWD slope decreases, which suggests that trees with higher leaf water content experience a smaller change in tree water deficit (TWD). This can be attributed to the role of leaf stored water during stress mitigation. Water storage, in all its forms, provides a mechanism to control the water status of leaf. Under water deficits, stored water may maintain plant viability. Buffering of stored water could serve as an advantageous drought response strategy (Leuschner et al., 2019; Meinzer et al., 2014).
  - H2: Visual stress signals such as leaf Chloris, bark injury or defoliation has a nonlinear effect on shrinkage. The "wiggly" effect is caused by stress levels that either enhance or decrease TWD. Long-term changes in water shortage and the use of the stored water are significantly influenced by species identification. Traits, visual stress signs from trees, and species explain a lot, but plot-level differences are also important. Overall, traits, visual stress signs from trees, species, and plots explain about two-thirds of why trees shrink differently.
- Report model fit statistics (e.g., R^2, conditional/marginal R^2).
  - H1: The conditional R^2 is 0.48, which explains the variance explained by the entire model (i.e. the fixed and random effects) and the marginal R^2 of 0.031 is the variance explained just by our fixed effect, which is a bit low too. This shows that our random effects explained a bit of our dataset. A higher conditional R^2 than a marginal R^2 simply means that the random effects explain additional variance compared to the fixed effects.
  - H2: The conditional R^2 from our mixed model where we accounted for the variance unaccounted for by plot location is 0.66, which explains the variance explained by the entire model (i.e. the fixed and random effects) and the marginal R^2 of 0.41 is the variance explained just by our fixed effects. This shows that our random effects explained a lot of our dataset. A higher conditional R^2 than a marginal R^2 simply means that the random effects explain additional variance compared to the fixed effects. The interclass correlation coefficient (ICC) is 0.42. This tells us that 42% of the unexplained

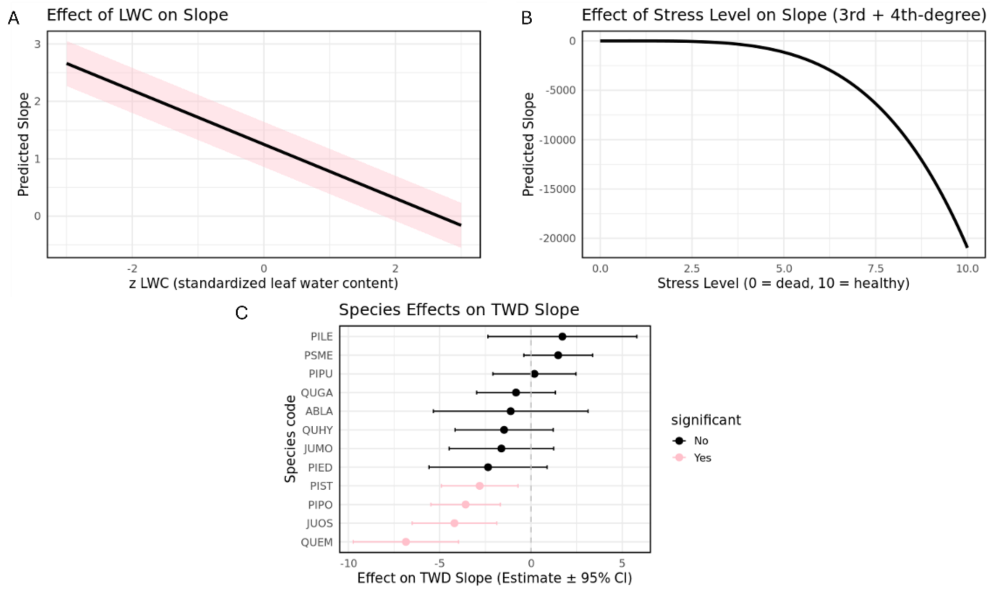
variance in our model is explained by our plot as random variables which is more than 40% of the variance in our response. So, while environmental variables and species difference is important, we should not underestimate the role of location, microclimate in the continuous drought mechanism.

5. **Create a Representative Figure of Your Model (5 points)**

- Include a figure that effectively communicates a key model result.
- Ensure the figure includes a descriptive caption, clear axis labels, and appropriate units.



- Examples: regression plot with CI, coefficient plot, or fitted vs. observed.
  - I fitted the response variable with the drop1() selected model (LWC) for H1.
  - For H2: I also used the drop1() model selection package and selected LWC, stress level and species are important predictors of TWD.



6. **Evaluate Model Assumptions (10 points)**

- Use residual plots, QQ plots, or other diagnostics to assess model assumptions.
- Discuss any violations and how you addressed them.
    - The initial model with all fixed effects and mixed effects didn't meet the linear model assumption, I did model selection to first drop irrelevant predictors and refit the model with the best selected model from model selection. After the model selection stage, I checked the model assumption again and the assumption improved.

7. **Appendix: Clean R Code (5 points)**

- Include clean, well-commented R code for all model fitting, outputs, and figure generation.

Submission Instructions

- Submit a single PDF file via Canvas
- Filename format: LASTNAME.FIRSTNAME.Assignment2.pdf