

Elemental Ecology
LDA Exercise
Due at Beginning of Class on Tuesday November 30th

The Marine Producer Library: Using multivariate techniques to find ecological patterns

For this assignment, you will be using the file "Producer_Library.csv". This file contains $\delta^{13}\text{C}$ data for 12 different amino acids from a wide variety of marine primary producers. These data were generated by Emma Elliott Smith as part of her dissertation investigating how different physiologies among marine algae may imprint on their isotopic compositions. In this assignment you will help Emma use multivariate statistical techniques to find differences, or similarities, among these taxa! *As always, remember to check that your working directory, the names of your files, and the names of the columns in the database match what you are telling R to look for!*

- 1) We want to see how well a linear discriminant analysis (LDA) can help us in distinguishing among red algae, green algae, kelps, and particulate organic matter (POM). To start, let's try an LDA using **all 12** amino acids. Refer to the code from the in-class demo to get started. **Address the following:**
 - i. Attach a pretty figure of producer amino acid $\delta^{13}\text{C}$ fingerprints (i.e., plot your producer groups along the first and second linear discriminant axes, LD_1 and LD_2 , and include a 95% confidence ellipse around each group).
 - ii. Which four amino acids are most informative for distinguishing among groups along LD_1 ? And, which four are the most informative for LD_2 ?
 - iii. What is the overall successful reclassification rate of this LDA model?
 - iv. Which groups are classifying well? Which are classifying poorly?

- 2) Now, let's pair down our inputs into the model, and run an LDA with only **five** amino acids.
 - i. Which amino acids did you use? Give a statistically, or biochemically, robust justification for why you chose these aminos.
Hint: Use your knowledge of amino acid classifications and your results from problem #1 to inform your decision.
 - ii. Provide a *brief* description of the biochemical affiliations for each of your chosen amino acids. Please include appropriate citations.
 - iii. Attach a new, pretty LDA figure!
 - iv. Does this smaller subset of data change your results from above? How? Be specific.
 - v. Are there any groups that are still not separating out from one another? Why do you think this might be? Provide a few citations that support your hypothesis.