

## **Regression Project: Progress Update**

Group 4: Elizabeth Burch, Kingsley Ofoegbu, Eli Schaub, Margaret Sun

Using Elizabeth's data we have begun work on building models to predict the presence of macroinvertebrates in seeps.

The 4 response variables we are considering are:

- Presence of *Stygobromus*
- Presence of *Crangonyx*
- Presence of any species
- Presence of multiple species.

Since our response variables are binary, we will be using a logistic regression for this project.

The list of potential explanatory variables is quite large and thus, to narrow this list, we have conducted a principal component analysis of the soil characteristics and we were able to reduce those 11 variables to just a couple of variables. Since we will be focusing on PCA for our other project, we will outline this process further within that project.

As noted above, the list of potential explanatory variables is large and should be narrowed, especially considering the smaller sample size. We have conducted both forward and backward stepwise regression to determine candidates for our regression models.

Other steps taken thus far include chi-square tests of both categorical and count variables as well as pie charts for the categorical variables.

Our next steps include checking model assumptions, asking further research questions, and compiling our results into a digestible presentation of our findings.