Biostatistics
Fall 2024

Final Capstone Project Feedback – 35 Points

Student: Andrew Bazuro Score: 30.75 /35 = 87.9 %

Part 1: Clean up your Repo – 5 pts Score: 5

Using Git/GitHub effectively and organizing a project well

Feedback: Github working well and project well organized and easy to understand.

Part 2: Finalize statistical analyses- 20 pts Score: 17

Remove unneeded code; Follow correct workflow; Reflects feedback; overall challenge

Feedback: Analysis 1 - when you create your model at line 137, if you don't specify a family, glm assumes gaussian, e.g. normal dist, e.g. you are doing a regular ANOVA and not accounting for normality. Otherwise pretty good.

Analysis 2 - Line 247 - again, you didn't specify a family. Looking at your histogram, you should likely be trying family = "poisson" as we did in class when we ran a glm. True for analysis 1 as well. You can see this in the output when it tells you that it used "gaussian".

Part 3: Final report – 10 pts Score: 8.75

Intro, Analysis with biological insight , Challenges; Well-written; Strong use of markdown

Feedback: Markdown looks good. Nothing fancy. I'd like to see you begin with intro, then setup and read in data for context. Imagine I'm a customer who paid you for this report - I'd want to start more big picture. You should explain what a seep is to give better biological context.

Writing could be a bit stronger. e.g. "The analysis focused on the question: What is the effect of different habitat types on Mg, Na, Fe, and K." you should be more clear - what is the effect of different habitat types on the concentration of several elements, Mg, Na, Fe, and K, in the source water" or something. Also, you need stronger biological/scientific justification for hypothesis, other than just citing a paper. Need to explain WHY you have a particular expected outcome.

Good job digging in on Tukey!