Biostatistics
Fall 2024

Final Capstone Project Feedback – 35 Points

Student: Sara St. Clair Score: 30 /35 = 85.7 %

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

Using Git/GitHub effectively and organizing a project well

Feedback: GitHub looks good. Project easy to navigate.

Part 2: Finalize statistical analyses- 20 pts Score: 16.5

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

Feedback: Q1 - Lifeform vs. biovigilance. Need to define what "biovigilance" actually means in terms of the data. You keep talking about "biovigilance categories" but it is a binary variable. Makes text confusing. You should save interpretation until after post-hoc analysis.Note - after post hoc you got a warning that you did not acknowledge or interpret. It is caused by having some groups with very small numbers. Appropriate procedure would be to combine some small groups together. As a result, I don't think your post hoc tests are reliable. Q2: Module vs. biogeo. Your chi-square is returning NaN and no p-values, likely because you cleared something that you needed when you reset from Q1. I'm taking off points because the code is not running correctly - good idea would have been to step through the final document. Also, no posthoc test needed if overall test is NS.
Q3 - Red_UK vs. lifeform. Same feedback for Q2.

Part 3: Final report – 10 pts Score: 8.5

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

Feedback: Markdown: some headings didn't render; good to figure out italics for Chisquare. Not much else fancy. In intro it might be nice to describe what some of the life form terms mean. Good description of data cleaning. I think you overstated how much you had to learn about post hoc testing in the challenges section, since I just showed you a package to run. This analysis was solid, but your data set lacked much in terms of interesting data types, so the analysis was fairly "vanilla" and thus it is even more important that it work correctly.