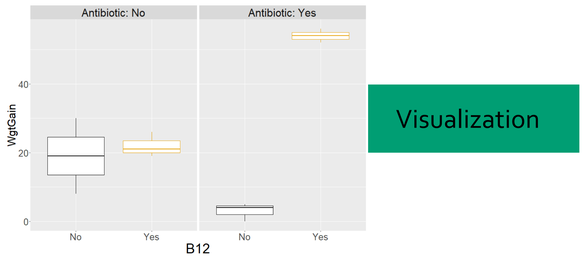
Exploring interaction terms in multiple linear regression using visualizations and fitted models

COLORADOSCHOOLOFMINES
EARTH • ENERGY • ENVIRONMENT

Daisy Philtron and Ashlyn Munson

Introduction

Interaction terms are consistently misinterpreted and misunderstood by students in 300 or 400 level linear modeling classes. How is understanding impacted when we teach using visualizations instead of only fitted models?



Visualization

Students in two sections were randomly divided to complete one of two activities. Both activities were asked the same questions about the same datasets.

- Activity A used **only** data visualizations
- Activity B using **only** fitted models

After the activity, all students took the same quiz on interaction terms. Then they switched activities and finished the learning.

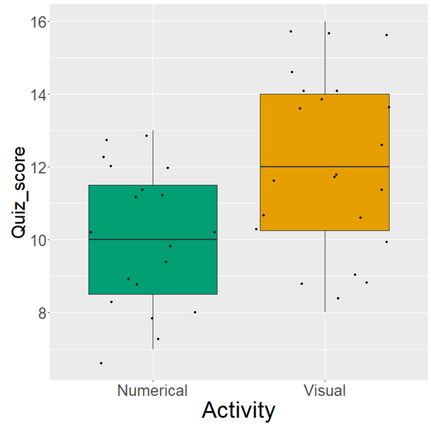
$$\widehat{Gain} = 19 - 16\times I(\text{Antibiotics=Yes}) + 3\times I(\text{B12 =Yes}) + 48\times I(\text{Antibiotics = Yes})\times I(\text{B12 = Yes})$$

Fitted model

Questions from the interaction quiz were included on a midterm and on the final.

Results

Post-activity quiz scores were significantly higher for the visual activity.

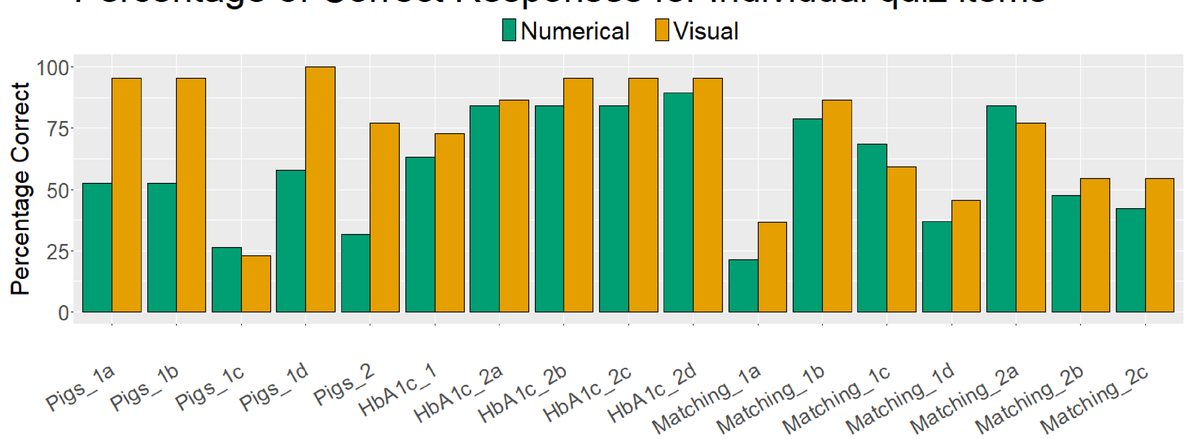


Difference in means: 2.2
t-stat: 3.30

Pigs 1 questions: T/F

a. In this study, adding antibiotics changes the effect of vitamin B12 with respect to weight gain.
b. The effect of adding B12 on weight gain is the same whether or not a pig is also taking antibiotics.
c. There appears to be a statistical relationship between B12 and Antibiotics.
d. The effect of adding B12 on weight gain is different depending on whether a pig is also taking antibiotics.

Percentage of Correct Responses for Individual quiz items



In later assessments, there was no significant difference between the groups.

Summary

- We recommend that professors incorporate visualizations, particularly during active learning, to teach the concept of interaction.
- Students performed significantly better on a post-test when they worked through an activity using only visualizations as opposed to only fitted models.
- The difference was largest for interaction terms related to categorical treatment effects.

Details

- The Colorado School of Mines is an engineering school with an emphasis on calculation and mathematical theory. All students take three semesters of calculus, for example.
- This experiment was performed in a 300-level statistics class with n = 19 students in the visual group and n = 22 students in the numerical group.
- Students that were not present in class for the activities were not included in any analyses.
- Most students in the class were Applied Mathematics, Statistics, or Data Science majors.
- For access to original activities, contact Daisy Philtron: dphiltron@mines.edu