

Module 8: Categorical Predictors

Jeffrey Woo

MSDS, University of Virginia

Welcome

- Remind me to record the live session!
- Recommended: put yourself on mute unless you want to speak.
- Reminder: the raise hand button can be found under “Manage Participants”.

Agenda

- Q&A
- Practical considerations with categorical predictors
- Small group discussion of guided question set
- Large group discussion of guided question set and other questions that popped up

Q&A

Terminology

- **Regression:** Quantitative response, predictors can be quantitative and/or categorical
- **Analysis of variance (ANOVA):** Quantitative response, categorical predictor(s).
- **Analysis of covariance (ANCOVA):** Quantitative response, categorical and quantitative predictor(s) (typically no interactions).

I personally view all of these through the regression framework, but some people do distinguish these.

Discrete Predictors

Do we treat discrete predictors as quantitative or categorical in our analysis? For visualizations:

- How many **distinct** values are there in the discrete variable?

Discrete Predictors

Do we treat discrete predictors as quantitative or categorical in our analysis? For models:

- Are we fine with assuming they have a “linear” relationship with the response variable?
- How many **distinct** values are there in the discrete variable?
- Are we concerned about needlessly adding parameters to our model especially if we have a small sample size?

Consideration of Interaction Terms

- Typically, people start with an **additive first order** model.
- Interactions considered at the start if:
 - Exploring interactions is part of your research question
 - An interaction makes sense contextually, or is well-established in the literature
 - You see evidence of interaction in your visualizations

Consideration of Interaction Terms

Can use some visuals to explore possibility of interactions (assuming quantitative response in all these scenarios):

- Interaction between a categorical predictor and a quantitative predictor: create scatterplot of response against quantitative predictor, create separate regression lines for each class.
- Interaction between two categorical predictors: side by side boxplots.
- Interaction between two quantitative predictors. If one is discrete, can use scatterplot approach, with separate regression lines for each value of the discrete variable. If both are continuous, you will have to create separate lines for various fixed values of one of the predictors.

Breakout Rooms

Upcoming

- HW 8 next Monday as usual.
- Group Evals for Mods 5 to 8 next Monday, via Canvas (link underneath the PDF for HW 8).
- New groupings for rest of semester (including project 2). Proj 2 opens next week.
- Module 9: model selection criteria and automated search procedures.