Module 7: General Linear F Test and Multicollinearity

Jeffrey Woo

MSDS, University of Virginia

Welcome

- You can download this set of slides. Find with the materials for the live session in Module 7.6.
- Remind me to record the live session!
- Recommended: put yourself on mute unless you want to speak.
- There is a "raise hand" button for you. Click on "Reactions" in the panel at the bottom.

Agenda

- Q&A
- Some comments on Module 7
- Small group discussion of guided question set
- Large group discussion of guided question set plus other questions that pop up

Q&A

Two main uses of regression models:

1. Prediction
2. Explore relationship between response and multiple predictors simultaneously.

- Including more predictors or higher order terms typically improves model fit (in fact R^2 never decreases), but also make the model more difficult to interpret.

- Making a model more complicated than needed can result in **overfitting**, which leads to poor predictive performance on new data.

Model Selection

- The general linear F test only works when comparing two models, the parameters of one being a subset of the parameters of another.
  - Some people call these nested models, but note that in experimental design, nested models refer to something else that is completely different.
- Other measures are used in other situations (module 9).

Solutions when Multicollinearity is Present

- Consider a subset of predictors (among those that are linearly dependent). Best approach if exploring relationships between predictor(s) and response. May help improve prediction on new data.

- Shrinkage methods. Should help improve prediction on new data but doesn't help in exploring relationships between predictor(s) and response.

- Dimension reduction methods. Should help improve prediction on new data. May help in exploring relationships between predictor(s) and response.

You will learn about the latter two methods in statistical learning.

Small Group Discussion

Module 8: Categorical predictors.

Due...

- Project 1, on Sunday, Mar 24. One upload per group. Submit via Assignments on Canvas.

- Project 1 Group Evaluation, on Sunday, Mar 24. Everyone completes. Submit via Assignments on Canvas.

- HW 7, on Monday, Mar 25, as usual.