SoDA 496 Review

Questions

- What is a data generating process?

- What is a model?

- What differentiates good models from bad models?

- What is a linear regression predicting?

- What is a generalized linear model?

- What is a virtual environment and why would you use one?

- What is validity? What are the different types of validity?

- What are the causes and consequences of omitted variable bias/measurement error/simultaneity bias/post treatment bias?

- Under what conditions should you expect these biases to occur?

- What is endogeneity and what are the implications of endogeneity?

- What are ways to alleviate endogeneity?

- What are fixed effects and why would you use them?

- What is the significance of random assignment?

- How can predictive modeling be used for statistical inference?

- In what ways can predictive modeling be used to alleviate some of the challenges of inferential modeling?

- What are the causes and consequences of overfitting?

- What is distributional shift and how does it relate to modeling?

Vocab

- Data Generating Process
- Model
- Inferential modeling
- Predictive modeling
- GLM
- Probability distribution
- Link function
- Normal, Poisson, Binomial, and Exponential distributions
- Virtual/Conda/Computing environment
- Dependency
- Construct validity
- Internal validity

- External validity
- Post-treatment bias
- Selection bias
- Endogeneity
- Omitted variable bias
- Measurement error
- Simultaneity
- Random assignment
- Fixed/Random effects
- Over/under fitting
- Out of sample prediction
- Distributional shift