Lesson 30

Clark

TEE Admin

Data Analysis Process

State a Research Question

Why do we build out statistical models? What are our objectives? Are we purely describing patterns in observed data? Are we formulating a model for the purpose of problem conceptualization? Are we examining a scientific theory? Are we making inference about a specific quantity? Are we predicting forecasting?

If prediction is our aim, we may choose to model data patterns that have no read explanation using structures we cannot interpret.

Explore the Data

What do we want to know about our *Random Variable*?

This might lead to a variety of different ways to explore our data. If our random variable is continuous what might we want to do?

If our random variable is binary, would that change how we approach our exploration?

What statistics would be beneficial to calculate at this stage?

Prepare the Data

What transformations might we consider?

What might we do with missing values?

What else might we want to do with our data?

Build and Interpret Models

Let's go through some of the models with their strengths and weaknesses.

Assess and Select the Best Model

Let's ensure we differentiate between model assessment and model selection. Model assessment is, does our model fit the data well. What might we look at here?

Model selection is, of a finite selection of models, which one will we use to represent our data. How do we go about accomplishing this?

Communicate the Results

Who are our stakeholders? What does it mean to conduct a stakeholder analysis? What are ethical considerations we might want to address?