

Lecture 1: Statistical Modeling  
POLSCI 702

Dave Armstrong  
UW Milwaukee

January 25, 2016

## Philosophy of Social Science Statistics

- Vauge (rather than mathematical) theories.
- Simple models → Complex reality.
- Fox - statistical models are almost always *descriptive*.
- Statistical models are not capturing social processes themselves.
  - Clarke and Primo suggest that models are like maps - they are objects (and thus cannot be true or false), have limited accuracy, represent only certain aspects of a real-world system.
  - Further, models are “purpose-relative” and model evaluation can only take place in the context of the purpose for which they were intended.

## What can Statistical Models tell Us?

- Describe the significant aspects of underlying structure in our data.
- Answer questions of practical/theoretical concern.
- Discover social facts that escaped initial consideration.
  - Particularly those that arise only when considering other things simultaneously, from conditional relationships or from other more complex patterns.

## Causal Interpretation

Occasionally, we may want to try to identify a causal effect. Here are three characteristics of our data and models that must be met before we can start to talk about causality.

1. Empirical relationship - there must be an empirical relationship between the variables for there to be a causal effect.
2. Temporal precedence - the cause needs to precede the effect in time.
3. Ruling out alternative explanations - all other plausible explanations of the phenomenon of interest must be ruled out (either through experimental or statistical control).

## Cautions for Statistical Modeling

- Recognize that models are mainly descriptive and attend to the descriptive accuracy of them.
  - We should not assume (without reflection) that relationships take a certain functional form, especially when the assumption is *very* strong and testable (like linearity).
- Be careful about using “causal” language - that one variable “affects” another is not to say that it is a “cause”.
- Use the appropriate statistical controls. Controls are variables that are not directly of interest, but:
  - influence the response variable, and
  - are related to the variable of interest.

## Example: Democracy and Repression

Let's consider some of my work as an example

- Cannot explain the “real process” whereby repression happens.
- Do statistics even make sense?
- Functional form assumption (does linearity make sense?)
- What can the model tell us?