Techniques for the Analysis of Legal and Judicial Data

2020 ICPSR Summer Program
13 July – 17 July

Professor: Miles T. Armaly
Email: milestarmaly@gmail.com OR mtarmaly@umich.edu

Course Description

This workshop introduces participants to a number of techniques that are useful in the analysis of legal and judicial data. Certain difficulties frequently plague scholars and practitioners in the analysis of such data, and this workshop offers a survey of techniques to handle these difficulties. Among the topics covered (among others) are analyzing necessarily nested data (e.g., justice votes on a case within a term), measuring latent traits (e.g., ideology), using text as data, and causal inference when traits (e.g., sex) cannot be randomized.

Course Prerequisites & Software

There are no prerequisites for this workshop. Ideally, however, participants will be comfortable with basic descriptive statistics and be familiar with linear models (e.g., ordinary least squares regression). Some basic understanding of maximum likelihood estimation will also be beneficial. The first day of the workshop will focus on OLS and MLE, should participants be unfamiliar with those models or need a refresher.

We will utilize the R statistical computing environment and RStudio, both of which are free. To remain focused on the substantive information, and not coding, all code to execute analyses will be available to participants.

Course Materials

There is no required text for this workshop. Because this workshop offers a survey of many techniques, the number of useful books would extend into the dozens. Below is a very, very short (and incomplete) list of resources that may be of use for topics covered in this workshop. Examples of applications are myriad across disciplines, and are omitted here. Sage has a series in Quantitative Applications in the Social Sciences (the “green books”). Many of the topics we will discuss are covered by this series. You may have access via your library.

- Long, J. Scott, and Jeremy Freese. Regression models for categorical dependent variables using Stata. Stata press, 2006/


Course Requirements

The primary requirement for this workshop is engagement with the course materials. There will be a handful of optional assignments intended to allow participants to apply (and troubleshoot) the methodologies that are covered in the workshop.

Class Schedule

The intention is to split the day into two parts and focus on one topic in the morning and one in the afternoon. Following each lesson, participants will have an opportunity to collaboratively complete assignments.

Given the nature of instruction – synchronous webcast – interruptions and technical difficulties are certain to occur. As such, it is difficult to know how much material will be covered in a given day. Additionally, should all participants be familiar with a given topic, it may be most prudent to move to the next topic. Thus, the schedule below is more of a chronological wishlist than a hard-and-fast schedule.

Monday July 13: Introduction & basics

Linear regression models (ordinary least squares)
Maximum likelihood estimation

Tuesday July 14: “Weird” standard errors

Clustered standard errors
Fixed and random effects
Multilevel models
Wednesday July 15: Measurement & dimensionality

Indexes, scaling, and reliability
Item response models
Factor analysis

Thursday July 16: Causal inference

Matching
Cross-lagged models

Friday July 17: Text as data & wrap-up

Topic modeling
Sewing together multiple topics