PSC 400

SYRACUSE UNIVERSITY

DATA ANALYTICS FOR POLITICAL SCIENCE

MULTIPLE REGRESSION

ASSIGNMENTS

- Review Exercise 6 due today
- Data Analysis Memo 3 due on Friday
- Will post Problem Set after class

CLASS TOPIC CHOICE

- Prediction (incl. loops)
- Geographic data (incl. how to make maps)
- Network data
- Text as data
- (Webscraping)

EXAMPLE

- social.csv
  - primary2006: 1 if voted, 0 if abstained
  - neighbors: 1 if received treatment, 0 if not
  - age: voter age in years

- How does the effect of the neighbors treatment vary by age?
  - Include age as well as interaction between neighbors and age in regression
  - Plot: predicted turnout by age when neighbors==1; and predicted turnout by age when neighbors==0

SURVEY

- **cces.csv**
  - age
  - female (1 if female, 0 otherwise)
  - nohighschool (1 if no high school degree, 0 otherwise)
  - collegeorhigher (1 if college attendance, 0 otherwise)
  - nonwhite (1 if not white, 0 if white)
  - married (1 if married, 0 otherwise)
  - employed (1 if employed full-time or part-time, 0 otherwise)
  - republican (1 if R, 0 if not)
  - democrat (1 of D, 0 if not)
  - impeach (1 if supports Trump impeachment, 0 if not)
  - votereg (1 if registered to vote, 0 otherwise)

SIDE NOTE

- Categorical variables
  - e.g. partisanship: Republican, Democrat, Independent
- Can be included as a set of binary (dummy) variables
  - republican (1 if R, 0 if not)
  - democrat (1 of D, 0 if not)
- Important: Have to leave out one category ("baseline")
  - Here: no dummy for Independents
  - Regression coefficients of R/D have to be interpreted relative to baseline

EXAMPLE

- What is the effect of age on whether respondents are registered to vote or not?