PS2030
Political Research and Analysis
Introduction to the Course

Spring 2025

WW Posvar Hall 3600

Professor Steven Finkel

Contact Info

- Steven Finkel

  4804 WW Posvar Hall

  Office Hours: Th 2:00-3:30 and by appointment

  E-mail: finkel@pitt.edu

  Web site: www.pitt.edu/~finkel

Goals of Course

- Further Your Knowledge of Basic Statistical Methods Used in Political and Social Science Research
  - Bivariate and multivariate regression
  - Assumptions underlying the "ordinary least squares" (OLS) methods for estimating regression parameters
  - Consequences of violations of the OLS assumptions
  - Variations on basic regression models: "dummy variables", non-linear, non-additive (interaction) models
  - Basic logit and probit models for dichotomous and ordinal dependent variables
  - Introduction to causal inference: regression, "matching" and propensity score analysis, endogeneity, instrumental variables, "difference-in differences", panel data methods, mediation analysis

- Improve your understanding and ability to critique existing quantitative political and social science research
- Develop your ability to conduct and present original statistical analysis relevant to your substantive interests
  - Three Homework Exercises (February 14, March 12, April 15)
  - Paper Assignment (April 18)
  - Poster Presentation (April 22)
- Further develop your skills with statistical software (Stata/R)
- Provide foundation for future courses in quantitative analysis
  - Time Series
  - Maximum Likelihood Estimation and Bayesian Methods
  - Longitudinal Analysis
  - Causal Inference and Experimental Methods

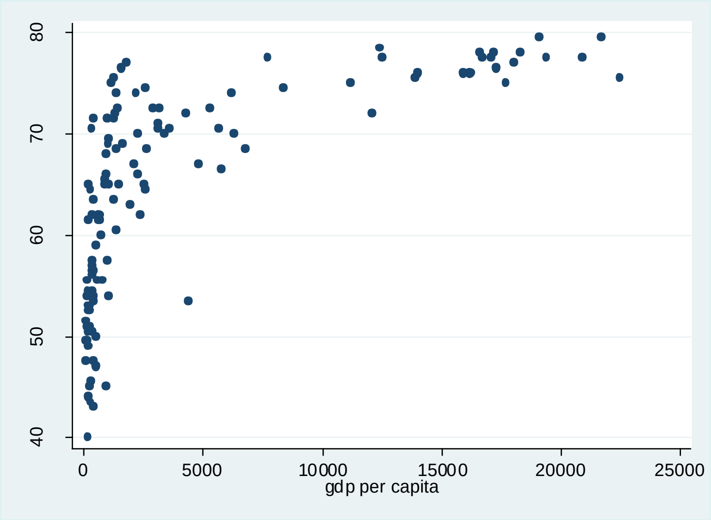
Overview of Course Material

- **Unit 1: Fundamentals of Linear Regression (weeks 1-4)**
- Bivariate Regression
  - Deterministic vs. probabalistic relationships
  - The population regression function
  - Estimating the "ordinary least squares" regression line from sample data
  - Properties of the OLS regression line
  - Interpreting the OLS intercept and slope coefficients
  - Interpreting the "goodness of fit" of the regression line

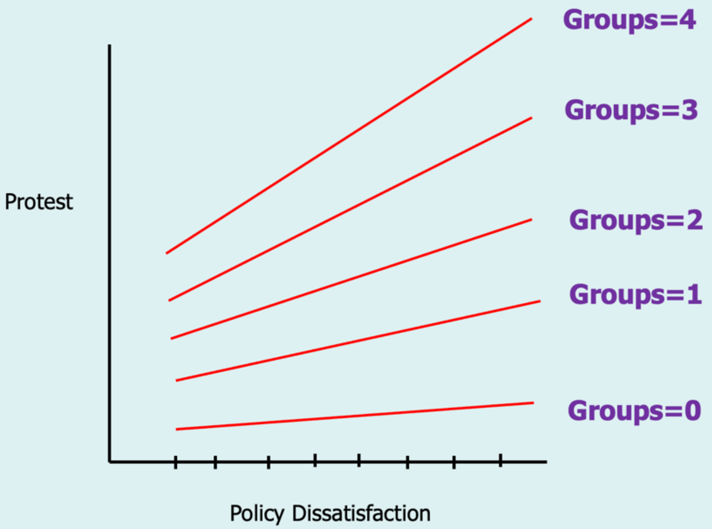
- **Unit 1: Fundamentals of Linear Regression (weeks 1-4)**
- Assumptions of Regression and Hypothesis Testing
  - Inferring population regression parameters from sample estimates
  - OLS *only* produces valid estimates of, and *only* permits valid inferences and hypothesis tests about population parameters to be made when certain assumptions about the population model hold What are they and why are they necessary?
  - Procedures for testing statistical significance of individual variables(t-test) and the model as a whole in terms of explanatory power (F test)
- Multiple Regression with more than one explanatory variable
- Regression with “dummy” (indicator) variables
- Mediation models

- **Unit 2: Regression Extensions and Problems (Weeks 5-7)**
- Nonlinear and Non-Additive Models
  - X may be related to Y in a *non-linear* fashion: As X increases, Y increases to a point and then decreases (or vice-versa)
  - X may relate to Y in a *non-additive* fashion as well, such that the effect of X_1 on Y depends on the level of X_2 . These are called "interactive" or "conditional effects" models and are very common

Nonlinear Example:
GDP (X) and Life Expectancy (Y)



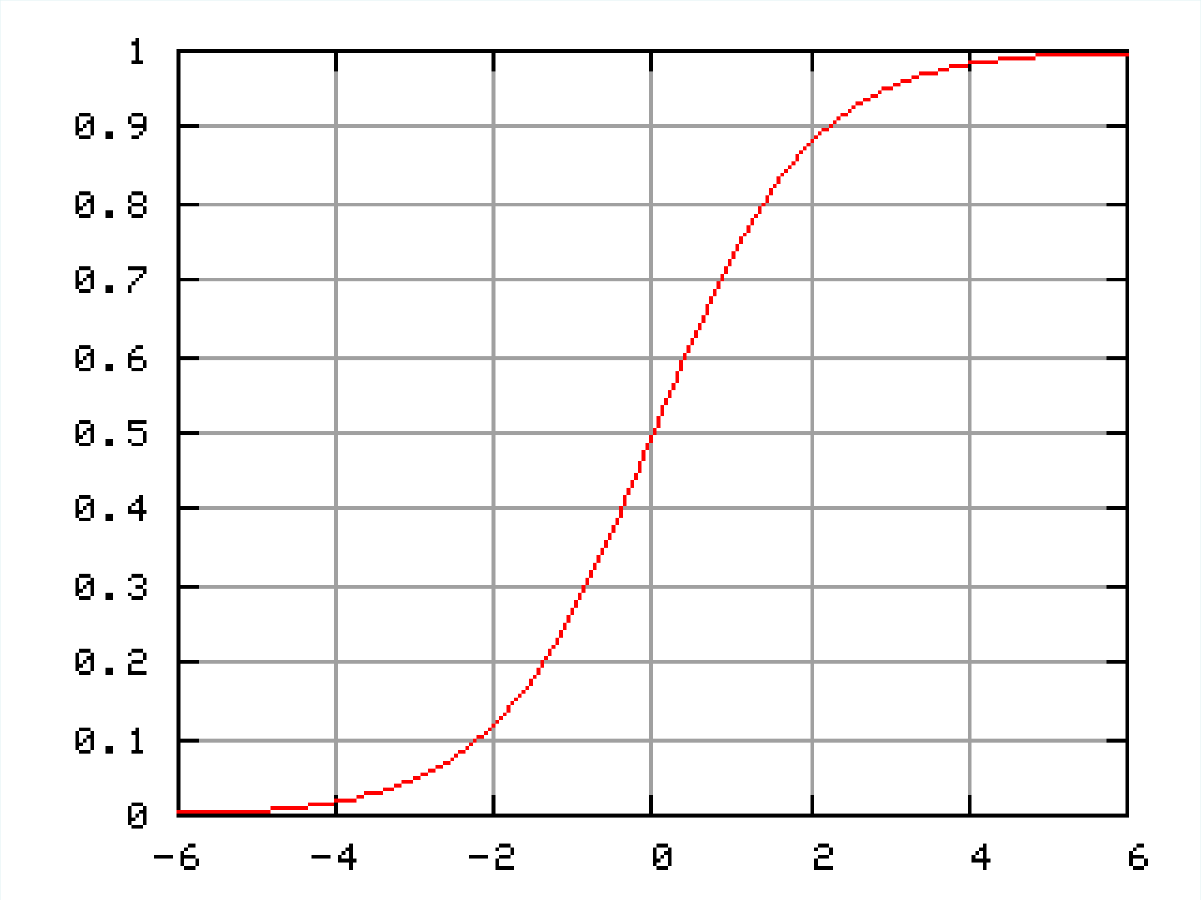
Nonadditive Example:
Protest (Y) as a function of Policy Dissatisfaction (X1) and Group Memberships (X2)



- Heteroskedasticity, Multicollinearity, Autocorrelation
  - Assumption of *homoskedasticity*, or constant error variance for all X.
  - Assumption that independent variables X_j and X_k are not *perfectly* related (*multicollinearity*)
  - Assumption that the error term for observation *i* is unrelated to the error term for observation *j* , (i.e. *no autocorrelation*) may not hold in the population for a variety of reasons (advanced topics covered in PS2740 Time Series Analysis)
- Endogeneous Regressors
  - OLS assumption of the independence of X and ε_i
  - Violation of this assumption indicates *endogeneity* in the model, and leads to serious problems with OLS estimation, inhibits ability to use OLS regression for making causal inferences about the effects of X on Y
  - Occurs when:
    - Important explanatory variables have been omitted from the model
    - Y causes X ("reverse causality") in addition (or instead of) X causing Y
    - X contains random measurement errors
  - Solutions: Instrumental Variables and Two-Stage Least Squares (and Unit 4)

- **Unit 3: Models for Non-Continuous Dependent Variables (weeks 8-10)**
  - Models thus far assume continuous dependent variables, but in practice many important DVs are dichotomous, ordinal, or multi-categorical
  - Dichotomous DVs (e.g., voted or abstained; war/no war) means we estimate the effect of X on the *probability* that Y takes on the value of 0 or 1
  - If we simply estimate a linear probability regression model, we will violate several OLS assumptions (heteroskedasticity, normality of errors, and possibly the assumption of linearity as well)
  - Solution: Transform the linear probability model into a non-linear model based on either the *logistic function* (i.e., the "logit" model) or the *cumulative normal function* (i.e., the "probit" model)

The Logistic Function



- Questions:
  - How do we estimate these models with Maximum Likelihood Methods, and how does this method compare with OLS?
  - How do we interpret logit or probit coefficients?
  - How do we report the "effects" of variables, given that the effect of a unit change in X differs at different points on the logistic (or cumulative normal) distribution?
  - How do we calculate and interpret goodness of fit measures in logit and probit models?
  - How can we extend these models to ordinal or nominal dependent variables, or other kinds of non-continuous outcomes (as a prelude to PS2730 Maximum Likelihood Estimation: Categorical and Limited Dependent Variables)?

- **Unit 4: Models for Causal Inference (weeks 11-13)**
  - Recent developments in causal analysis see regression as one of several important tools for estimating causal effects
  - We will discuss the basics of the modern approach to causality, the "potential outcomes" or "counterfactual model", and then discuss the role of regression and other methods such as:
    - Randomized experiments
    - "Matching methods", where observations at different levels of X are "matched" or balanced on other observed Z factors that may affect Y before calculating the causal effect of X. (e.g., "propensity score matching")
    - Advanced instrumental variable analysis to identify specific kinds of causal effects in experimental and in observational studies
    - "Difference-in-differences" and panel models for longitudinal data (as a prelude to PS2701 Longitudinal Analysis)
    - All as prelude to PS2702 advanced methods course in Causal Inference

Your Professor

- Academic Positions
  - Daniel Wallace Professor of Political Science, University of Pittsburgh, 2005—present
  - Department Chair, 2011-2018
  - Professor of Quantitative and Qualitative Methods, Hertie School of Governance, Berlin, Germany, 2005 – 2008
  - University of Virginia, 1984-2005
  - PhD Stony Brook University (year not disclosed!)
- Research Interests
  - Political Behavior, Participation, Democratization
  - Civic Education and the Development of Democratic Political Culture
  - Evaluation of Democracy Assistance Programs of USAID and International Donors
  - Statistical Methods for Longitudinal and Panel Data
- Home Page: www.pitt.edu/~finkel