Segment 1: Introduction
Prof. Sharyn O’Halloran
Preliminaries

- Professor Sharyn O’Halloran
- Office Hours: Wednesday 9:00 to 11:00
- Contact information:
  - Office: 727 IAB
  - E-MAIL: SO33@columbia.edu
  - PHONE: X4-3242
Overview

- How to **describe** numerical data.
- How to make **inferences** about populations from samples.
- How to **evaluate** the relation between variables, factors or events.
Example 1: Education

- **Main Finding:**
  - Black and Hispanic students are far less likely to attend college than are white students.
Example 1: Education (cont.)

- **Dependent Variable:**
  - What is the phenomenon to be explained?
  - Here, it is the percent of minorities enrolled in college

- **Independent Variable:**
  - What factors might help explain this phenomenon?
  - We can think of many...
Example 1: Education (cont.)

- **Evidence:**
  - 34% of Whites 18-24 were enrolled in college in 1991.
  - 24% of Blacks 18-24 were enrolled in college in 1991.
  - 18% of Hispanic 18-24 were enrolled in college in 1991.
Example 1: Education (cont.)

- How would you represent this data graphically?

Percent Attending College by Category
Example 1: Education (cont.)

- Measurement Problems
  - What does 34% represent?
  - What is the reference group?
  - What is the sample population?
What is the causal relation between race & education?

- Education is the dependent variable or the thing to be explained.
- Race is the independent variable or the causing factor.

This is a causal model or a path diagram.

Simplest model:

Race  →  Education

(Independent)  (Dependent)
Example 1: Education (cont.)

- What does the article suggests?
  - Race → Income → Education
    - Independent → Intervening → Dependent

- Income is an **intervening** variable.
  - Because minorities tend to have lower incomes, they are less able to afford education.

- Implication
  - Race affects education via income
Policy Prescription:
- Article argues that to improve educational attainment, need to ensure funding for minorities.

But what if income is not the problem?
- What if the relation between race and higher education is due to discrimination or cultural factors?
- How should we redirect government policy?

Example 1: Education (cont.)
A second hypothesis postulated is that:

- Race → Income → Dropouts
  - (Independent) → (Intervening) → (Dependent)

Example 1: Education (cont.)
Policy Implication:
- Raising the minimum wage will increase dropout rates.

Moral:
- Different models of the world lead to different policy predictions.
Example 2: Environment

- What other factors might intervene here?
- How would these interpretations change the implied policy prescriptions?
http://www.harpers.org/harpers-index/listing.php3

Numbers are present as facts, as if they speak for themselves

But numbers rarely speak for themselves
  - As we have seen, they can have many different interpretations and causes

This course will teach you how to speak for the numbers (or else someone will do it for you)
Goals of the Course

- The purpose of the course is to introduce professional students to basic data analysis skills.
  - Develop techniques to test and evaluate competing models of how the world works.

- Approach
  - Hands on / learning by doing
Materials

- **Text**
  - Wonnacott and Wonnacott (4th Edition)
  - Course Packet

- **Software**
  - SPSS for Windows (Also available at the CU Bookstore)
  - Excel
    - SIPA Skills Course

- **Data**
  - 1998 GSS Data set available on the SIPA server
  - See “Why Take Statistics” in Course Packet
Teaching Assistants

- One TA
  - Head Sections and Office Hours
  - Weekly Labs Mandatory
    - Meet in classroom, then move to lab

- One PRA
  - Grading
  - Office Hours
Support

- **Website**
  - [http://www.columbia.edu/itc/sipa/U4320y-003](http://www.columbia.edu/itc/sipa/U4320y-003)

- **Newsgroup**
  - Class bulletin board
    - Check regularly
    - Be-Nice Policy

- **Class Notes**
  - PowerPoint slides available on website after class
  - Not a substitute for attending lecture
Grading

- Weekly Assignments (40%)
  - No late work
  - Presentation counts

- Midterm (30%)
  - In class
  - 1 3x5 index card

- Final Paper (30%)
  - Can work in pairs