Sociology 30005
Statistical Methods of Research 2
Assignment 1

Probability

Consider the population of US adults in the labor force in 1992. We are interested in the relationship between educational attainment and the risk of unemployment. Educational attainment has six possible values: no degree, GED, high school degree, associates degree, bachelors degree, and masters degree or higher.

1. Construct a theoretical contingency table with two rows (values of unemployment) and six columns (values of educational attainment) in which the entries are the joint and marginal probabilities. Use Greek letters to represent these (for example use π).

2. Define the marginal probability of unemployment and decompose it into the sum of the relevant joint probabilities.

3. For each possible level of education, define the conditional probability of unemployment.

4. Decompose the joint probability of having no degree and being unemployed into the relevant marginal and conditional probabilities.

5. Now decompose the marginal probability of being unemployed into a function of the relevant marginal and conditional probabilities.

6. Using the NALS data, estimate the conditional probabilities of unemployment for each level of education. What does this seem to say about the association between education and unemployment?

7. Again using NALS, assume that unemployment and education were independent. What would then be the estimated conditional probability of unemployment given no degree? Under this scenario, how many of those with degree would we expect to be unemployed? Compare this to the number of those with no degree who were in fact unemployed and comment on how education is associated with joblessness for this group.

Expectation

We are again going to work with NALS, now using three variables: parent years of education (X), respondent years of education (Z), and adult literacy (Y).

1. Write down a theoretical linear regression model in which Y is a function of X and Z. Define the terms in the model. We will assume this is the "true model" of the relationship between X,Z, and Y.

2. Write down two other linear regression models: a) Z is a function of X; and b) X is a function of Z;

3. Now suppose someone estimated a model using only Z as a predictor. That means the person would be studying the expected value of Y given Z alone.

   a. Using (1), find E(Y|Z).
   b. Using (1) and (2b), define the bias involved. Show that it has two parts and define them.

4. Now suppose someone estimated a model using only X as a predictor. That means this person would be studying the expected value of Y given X alone (2c).

   a. Using (1), find E(Y|X). Define the "total effect" of X on Y.
   b. What is the direct effect of X on Y based on your theoretical model (1)?
   c. Find the indirect effect of X on Y as it operates through Z.
   d. Show that the total effect of X on Y is the sum of the direct and indirect effects you have defined.

5. Estimate these total, direct, and indirect effects using NALS, and comment on what you have learned about how parent education and respondent education are linked to adult literacy.