Econometrics for Public Policy II
Spring 2020

Problem Set 1
Due Class 3 (January 29) in class on paper

Please find the datasets for this problem set under the "handouts" section of the website. There are two datasets: one for 1950 and another for 2010.

Each dataset has one observation per US county in that year (1950 or 2010). Data come from the Decennial Cenus (1950, some 2010) and the American Community Survey (2010, which is really 2008-2012 5-year average). All variables are labeled. The census tabulates data from the individual collection at a variety of levels of geography; here we are using county-level data.

The variables statefips/countyfips uniquely identify observations in each dataset. You can find definitions for the statefips and countyfips variables at `https://census.missouri.edu/geocodes/` and many other websites.

Please turn in a set of written (preferably typed) answers to these problems, as well as a do file. The do file should have comments that indicate the commands associated with each question.

1. Summary statistics

   a. By year, find the average across all counties for

      - population
      - log of population
      - share white
      - share black
      - share women age 25+ with education of some college or more
      - share men age 25+ with education of some college or more

   b. Find averages of the same variables by year and state

2. Matching Data

   a. How many counties are in both the 1950 and 2010 datasets?

   b. How many counties are in the 1950 dataset, but not the 2010 dataset?

   c. How many counties are in the 2010 dataset, but not the 1950 dataset?

d. Investigate two counties that are in the 2010 dataset, but not the 1950 dataset. Why is this?

e. Investigate two counties that are in the 1950 dataset, but not the 2010 dataset. Why is this? (This one is trickier! If you get stuck here, don't spend tons of time on this.)

3. Regressions

a. Make a panel dataset from 1950 and 2010 (you may have already found this useful for the previous question)

b. Regress log of population on the four share variables you created above

c. Interpret one of the share coefficients

d. Repeat the previous regression with state fixed effects

e. Interpret one of the share coefficients with the inclusion of the state fixed effects

f. Report how much a one standard deviation change in your share of interest impacts population