Econometrics for Public Policy II
Fall 2023

Problem Set 1
Due before Class 3, September 13

There are two datasets for this problem set: one for 1950 and another for 2010. (You should see links in the previous sentence for data downloading.) If you'd like non-Stata datasets, please let me know and I'm happy to provide.

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 use county-level data.

The variables statefips/countyfips uniquely identify observations in each dataset. You can find definitions for the statefips and countyfips variables at this helpful webpage from the University of Missouri. and many other websites.

Use Stata's `describe` command to see the definitions of the remaining variables.

Please turn in a set of written answers to these problems, as well as a do file (or program from the statistical software of your choice). The program file should have comments that indicate the commands associated with each question.

You are both welcome and encouraged to work on this problem set with your classmates. The problem set you turn in should be your own work – both the code and the written output. If we notice exactly duplicative work, we will give zero credit to both assignments.

1. Summary statistics

   a. Make a panel dataset from 1950 and 2010, meaning a dataset that has one observation per county and year. In this dataset, most counties will have two observations, one for 1950 and one for 2010. Stata's `append` command stacks one dataset on top of another.

   b. By year, find the average of

      - population (`cv1`)
      - log of population (create yourself from `cv1`)
      - share white (`s1`)
      - share black (`s2`)
      - share women age 25+ with education of some college or more (`s3`)

- share men age 25+ with education of some college or more (`s4`)

Use one command to find all these averages. In Stata, you can use `collapse` combined with `, by(year)`.

After using `collapse`, you can use `outsheet` to output the resulting dataset as a txt or csv. Using this output file, it should not be difficult to create a labeled table.

Answer: See Table 1 at end.

c. Find averages of the same variables by year and state for California, Mississippi and New Jersey.

Note that California's state code is 6, and that it has a leading zero – so write it `06`. Again, use Stata's `collapse` combined with `, by(year state)` to generate these outputs. Also again, you can use Stata's `outsheet` command to output the data you've created to a .csv or .txt file.

Answer: See Table 2 at end.

2. Matching Data

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

In the previous question you created a panel dataset. To answer this question, you may prefer to make a "wide" dataset with one observation per county. It may be helpful to make a indicator variable (0/1) for having an observation in a given year in the 1950 and 2010 datasets, and merge the datasets.

In the merged dataset, you can use Stata's `tab` command to see a cross-tab of two indicator variables. For example, if your variables are called `y1950` and `y2010`, you can tell Stata to report `tab y1950 y2010`. Correctly interpreting this table will tell you the answer to the first three parts of this question.

Answer: 3,090

See notes in program and log file about how to find this.

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

Answer: 12

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

Answer: 19

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

Answer: Here are two examples – you answer can include any valid examples. My two examples did not exist in 1950.

- Menominee County, Wisconson (55/078) was created in 1959 (see Wikipedia)
- La Paz County, Arizona (04/012) was established in 1983 (again, see Wikipedia)

3. Regressions

a. Return to the panel dataset from question 1.

b. Regress log of population on the four share variables you created in question 1 and a fixed effect for year = 2010.

For this and the next question, it is sufficient to paste the results from the log; for future problem sets you will need to make a regression table, but you do not need one here.

In Stata, there are multiple says to create indicator variables and use them in a regression. Here are two equivalent methods:

- ```
  gen y2010 == 0
  replace y2010 = 1 if year == 2010
  regress y x y2010
  ```
- `xi: regress y x i.y2010`
- You can test for yourself whether these yield equivalent results

Answer: Results are in the log file.

c. Interpret the coefficient on the year indicator variable

Answer: The coefficient on the year indicator I estimated is -0.605, which means that the average county has 0.605 log points lower population in 2010 than in 1950. This negative coefficient may strike you as surprising, but remember that the US has had a big shift to urban areas – so while the biggest counties got bigger, most counties lost population.

Usually when the dependent variable is in logs, we can interpret the coefficient as a percentage change. We can do this because the coefficient tells us that there is a β log point change in the dependent variable for a one-unit change in X. To convert this log point change into a regular old change, we do e^{β}, since $\beta = \ln(\Delta Y)$. Exponentiating both sides gives $e^{\beta} = e^{\ln(\Delta Y)}$, or $e^{\beta} = \Delta Y$, where Δ denotes the change.

For small β, $e^{\beta} \sim 1 + \beta$, so we can interpret, for example, $\beta = 0.03$ as a 3% change. However, our change is pretty big! So let's do the math: $e^{-0.606} = 0.546$, or a 55%
  ```

decline, rather than the 60% decline the naive interpretation of the coefficient would suggest.

For this question, I was hoping that you would interpret the coefficient as an average decline in population of about 60%.

d. Repeat the previous regression with state fixed effects

Answer: Results are in the log file.

e. Interpret one of the share coefficients from the second regression

Answer: First note that the shares and percentages are equivalent. A share of 0.01 is 1 percent. A share of 1 is 100 percent. A one unit change in the share is a change from 0 to 1, which is a change from 0 percentage points to 100 percentage points.

In the regression, the coefficients are -2.06 (share white), -1.02 (share AA), -3.8 (share women at least college) and 9.00 (share men at least college). Using the first one, a one hundred percentage point increase in the share of the white population (a one-unit change) is associated with a 206 percent decrease in a county's population. (Because the dependent variable is in logs, we can interpret the coefficient as a percentage point change.)

However, no counties experience a one hundred percentage point decline in white population share.

4. Long and Short Regressions and Omitted Variable Bias

- From the lecture, we learned the omitted variable bias formula. Now you're going to calculate a specific example.
- We limit our analysis just to 2010.
- We are interested in the impact of the share of college educated men on the employment to population ratio and on the extent of omitted variable bias if we exclude the share of women who are college educated.
  - Let E_i, defined as `cv59 / cv1`, denote the employment to population ratio in county i
  - Let M_i be the share of men age 25 or above who are college educated in county i
  - Let W_i be the share of women age 25 or above who are college educated in county i
- Let's suppose that we have a "true" long equation

$$E_i = \beta_0 + \beta_l M_i + \gamma W_i + \epsilon_{l,i} \tag{1}$$

- However, we sometimes want to be lazy and estimate a "short" regression:

$$E_i = \beta_0 + \beta_s M_i + \epsilon_{s,i} \tag{2}$$

- How bad is the short regression? The omitted variable bias formula tells us that

$$\beta_s - \beta_l = \pi * \gamma \tag{3}$$

  where γ is the coefficient on W from the long regression and π is the coefficient on M from this regression that estimates the strength of the correlation between M and W:

$$W_i = \alpha + \pi M_i + \epsilon_{c,i} \tag{4}$$

a. Estimate equations 1, 2 and 4 above.

b. Use your estimated coefficients to show that the omitted variable bias formula (equation 3) holds. To do so, write the estimates for β_s and β_l, and show that their difference is equal to the difference between your estimates of π and γ.
Answer: Results are in the log file.

How to Turn This In

See the syllabus for how to create your GW Box folder. Send an invitation to this folder to me and our TA Genevieve. You'll turn in all your work to this folder.

For this assignment, create a sub-folder called "Problem Set 1". In this folder, you should upload two files. The first is the responses to these questions. Name this file "problem_set_1_lastname.pdf" (pdf preferred). The second file is the code that generates these results **in pdf format**. Save this file as "problem_set_1_code_lastname.pdf"

Table 1: National County Averages by Year

	1950	2010
population	48581	98641
log(population)	9.9	10.3
Share white	0.891	0.843
Share African American	0.101	0.09
Share of women age 25+ with at least some college	0.119	0.515
Share of men age 25+ with at least some college	0.1	0.466

Table 2: State Means by Year

				share			
						at least college, age 25+	
State	year	population	log(pop)	white	African Am.	women	men
CA	1950	182521	10.8	0.951	0.019	0.168	0.157
	2010	642310	12	0.752	0.033	0.613	0.575
MS	1950	26572	10	0.564	0.435	0.086	0.075
	2010	36187	10.1	0.565	0.41	0.477	0.41
NJ	1950	230254	11.9	0.936	0.063	0.106	0.138
	2010	418662	12.7	0.738	0.121	0.575	0.573

```
# delimit;

*************************************************************************

this problem set asks students to do the following things
 - take average by year (stack data)
 - merge data across years (figure out which obs arent consistent across years)
 - collapse by year and state
 - do a simple regression
 - find average change in share educated by state

january 17, 2017
february 14, 2017
january 16, 2018
february 11, 2020 ** update to fix
january 11, 2022
august 30, 2023
august 31, 2023

ps1v04.do

*************************************************************************;

clear all;
pause on;
set more off;

capture log close;
log using ps1.log, replace;

dateo;

***** 0. prepare data for students ***************************************;

*** bring in data ***;

* bring in 1950 data *;
use /home/lfbrooks/pppa6022/2017/problem_sets/stata_basics/data/d1950_20170117;

* append (stack) 2010 data *;
append using
/home/lfbrooks/pppa6022/2017/problem_sets/stata_basics/data/d2010_20170117;

* dataset should be unique by statefips/countyfips/year *;
duplicates report statefips countyfips year, analyze;

*** calculate needed variables ***;

* share white *;
gen s1 = (cv1 - cv3 - cv4) / cv1 ;
```

```
label variable s1 "share white";

* share black *;
gen s2 = cv3 / cv1;
label variable s2 "share black";

* share of women age 25+ college educated *;
gen s3 = (cv25 + cv26)/(cv18 + cv19 + cv20 + cv21 + cv22 + cv23 + cv24 + cv25 +
cv26) if year == 1950;
replace s3 = (cv25 + cv26 + cv27)/(cv18 + cv19 + cv20 + cv21 + cv22 + cv23 + cv24 +
cv25 + cv26 + cv27) if year == 2010;
label variable s3 "share of women age 25+ college educated";

* share of men age 25+ college educated *;
gen s4 = (cv15 + cv16)/(cv8 + cv9 + cv10 + cv11 + cv12 + cv13 + cv14 + cv15 + cv16)
if year == 1950;
replace s4 = (cv15 + cv16 + cv17)/(cv8 + cv9 + cv10 + cv11 + cv12 + cv13 + cv14 +
cv15 + cv16 + cv17) if year == 2010;
label variable s4 "share of men age 25+ college educated";

* keep needed variables *;
keep statefips countyfips year cv1 cv59 s1 s2 s3 s4 d1950 d2010 name;

* preserve so I can save a 1950 and a 2010 *;
preserve;

* save 1950 version *;
keep if year == 1950;
drop d2010;
save /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_${date}, replace;

* save 2010 version *;
restore;
keep if year == 2010;
drop name d1950;
save /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_${date}, replace;

***** 1. averages by year, and by state and year ***************************;

* find average share white, average share black,
  average share any college or more by gender
  in both years (national)
  and by state (make output a dataset, not just printed to the screen)*;

*** bring in data ***;

* date of data *;
local date_of_data "20230830";

* bring in 1950 data *;
```

```
use /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_`date_of_data';

* append (stack) 2010 data *;
append using
/home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_`date_of_data';

* dataset should be unique by statefips/countyfips/year *;
duplicates report statefips countyfips year, analyze;

*** calculate needed variables ***;

* log of population *;
gen ln_cv1 = log(cv1);
label variable ln_cv1 "log of population (cv1)";

* save for later use *;
save /groups/brooksgrp/junk/uselater, replace;

*** find national averages by year ***;

preserve;
sort year;
collapse (mean) cv1 ln_cv1 s1 s2 s3 s4, by(year);
outsheet using output/natl_lvl_averages.txt, replace;

list;

*** find averages by state and year for CA, MS and ME ***;

restore;
preserve;
keep if statefips == "06" | statefips == "28" | statefips == "34";
table statefips;
sort statefips year;
collapse (mean) cv1 ln_cv1 s1 s2 s3 s4, by(statefips year);
outsheet using output/state_lvl_averages.txt, replace;


****** 3. county change over time
*******************************************************;

* clear all data *;
drop _all;

* bring in 1950 data *;
use /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_`date_of_data';
keep statefips countyfips d1950 cv1 name;
rename cv1 cv1_1950;
rename name name_1950;
sort statefips countyfips;
```

```
* merge in 2010 data *;

merge 1:1 statefips countyfips using
  /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_`date_of_data';

* fix markers so i can clearly see merging issues *;
replace d1950 = 0 if d1950 == .;
replace d2010 = 0 if d2010 == .;

* look at cross-tab of which counties appear when *;
tab d1950 d2010;

* list the counties that dont merge *;
list statefips countyfips year d1950 d2010 name_1950 name if d1950 + d2010 != 2;


****** 4. regression
**********************************************************************;

* 3(a) *;
*** bring in data from 1 that I saved ***;
drop _all;
use  /groups/brooksgrp/junk/uselater;

* 3(b) *;
* population as a function of our four variables, with and without state fixed
effects *;
regress ln_cv1 s1 s2 s3 s4;
regress ln_cv1 s1 s2 s3 s4 i.year;

* 3(c) *;
* answer in words *;

* 3(d) *;
xi: regress ln_cv1 s1 s2 s3 s4 i.statefips;
xi: regress ln_cv1 s1 s2 s3 s4 i.statefips i.year;

* 3(e) *;
* answer in words *;

* 3(f) *;
* find the standard deviation in share white *;
summ s1 if year == 2010;
* answer in words *;

***** 4. long and short regressions ****************************;

* look at 2010 employment-to-population ratio as a function of male and female
education **;
```

```
-------------------------------------------------------------------------------
-----------------------------------------
      name:  <unnamed>
       log:  /home/lfbrooks/pppa6022/2023/problem_sets/ps1/ps1.log
  log type:  text
 opened on:  30 Aug 2023, 10:44:48

. dateo;

. ***** 0. prepare data for students ****************************************;
. *** bring in data ***;
. * bring in 1950 data *;
. use /home/lfbrooks/pppa6022/2017/problem_sets/stata_basics/data/d1950_20170117;

. * append (stack) 2010 data *;
. append using
/home/lfbrooks/pppa6022/2017/problem_sets/stata_basics/data/d2010_20170117;
(note: variable cv87 was int, now double to accommodate using data's values)
(note: variable cv11 was long, now double to accommodate using data's values)
(note: variable cv12 was long, now double to accommodate using data's values)
(note: variable cv21 was long, now double to accommodate using data's values)
(note: variable cv22 was long, now double to accommodate using data's values)

. * dataset should be unique by statefips/countyfips/year *;
. duplicates report statefips countyfips year, analyze;

Duplicates in terms of statefips countyfips year

--------------------------------------
   copies | observations       surplus
----------+---------------------------
        1 |         6211             0
--------------------------------------

. *** calculate needed variables ***;
. * share white *;
. gen s1 = (cv1 - cv3 - cv4) / cv1 ;

. label variable s1 "share white";

. * share black *;
. gen s2 = cv3 / cv1;

. label variable s2 "share black";

. * share of women age 25+ college educated *;
. gen s3 = (cv25 + cv26)/(cv18 + cv19 + cv20 + cv21 + cv22 + cv23 + cv24 + cv25 +
cv26) if year == 1950;
(3,109 missing values generated)
```

```
. replace s3 = (cv25 + cv26 + cv27)/(cv18 + cv19 + cv20 + cv21 + cv22 + cv23 + cv24
+ cv25 + cv26 + cv27) if year == 2010;
(3,109 real changes made)

. label variable s3 "share of women age 25+ college educated";

. * share of men age 25+ college educated *;
. gen s4 = (cv15 + cv16)/(cv8 + cv9 + cv10 + cv11 + cv12 + cv13 + cv14 + cv15 +
cv16) if year == 1950;
(3,109 missing values generated)

. replace s4 = (cv15 + cv16 + cv17)/(cv8 + cv9 + cv10 + cv11 + cv12 + cv13 + cv14 +
cv15 + cv16 + cv17) if year == 2010;
(3,109 real changes made)

. label variable s4 "share of men age 25+ college educated";

. * keep needed variables *;
. keep statefips countyfips year cv1 cv59 s1 s2 s3 s4 d1950 d2010 name;

. * preserve so I can save a 1950 and a 2010 *;
. preserve;

. * save 1950 version *;
. keep if year == 1950;
(3,109 observations deleted)

. drop d2010;

. save /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_${date}, replace;
file /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_20230830.dta saved

. * save 2010 version *;
. restore;

. keep if year == 2010;
(3,102 observations deleted)

. drop name d1950;

. save /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_${date}, replace;
file /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_20230830.dta saved

. ***** 1. averages by year, and by state and year **************************;
. * find average share white, average share black,
>   average share any college or more by gender
>   in both years (national)
>   and by state (make output a dataset, not just printed to the screen)*;
. *** bring in data ***;
. * date of data *;
```

```
. local date_of_data "20230830";

. * bring in 1950 data *;
. use /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_`date_of_data';

. * append (stack) 2010 data *;
. append using
/home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_`date_of_data';

. * dataset should be unique by statefips/countyfips/year *;
. duplicates report statefips countyfips year, analyze;

Duplicates in terms of statefips countyfips year

--------------------------------------
   copies | observations       surplus
----------+---------------------------
        1 |         6211             0
--------------------------------------

. *** calculate needed variables ***;
. * log of population *;
. gen ln_cv1 = log(cv1);

. label variable ln_cv1 "log of population (cv1)";

. * save for later use *;
. save /groups/brooksgrp/junk/uselater, replace;
file /groups/brooksgrp/junk/uselater.dta saved

. *** find national averages by year ***;
. preserve;

. sort year;

. collapse (mean) cv1 ln_cv1 s1 s2 s3 s4, by(year);

. outsheet using output/natl_lvl_averages.txt, replace;

. list;

     +-------------------------------------------------------------------------------+
     | year           cv1     ln_cv1         s1         s2         s3         s4 |
     |-------------------------------------------------------------------------------|
  1. | 1950   48580.70954   9.894452   .8910063   .1009891   .1189899   .1004977 |
  2. | 2010   98641.04407   10.27845    .842963   .0904683   .5153075   .4664086 |
     +-------------------------------------------------------------------------------+

. *** find averages by state and year for CA, MS and ME ***;
. restore;
```

```
. preserve;

. keep if statefips == "06" | statefips == "28" | statefips == "34";
(5,889 observations deleted)

. table statefips;

----------------------
state     |
fips code |      Freq.
----------+-----------
       06 |        116
       28 |        164
       34 |         42
----------------------

. sort statefips year;

. collapse (mean) cv1 ln_cv1 s1 s2 s3 s4, by(statefips year);

. outsheet using output/state_lvl_averages.txt, replace;

. ****** 3. county change over time
*********************************************************;
. * clear all data *;
. drop _all;

. * bring in 1950 data *;
. use /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d1950_`date_of_data';

. keep statefips countyfips d1950 cv1 name;

. rename cv1 cv1_1950;

. rename name name_1950;

. sort statefips countyfips;

. * merge in 2010 data *;
. merge 1:1 statefips countyfips using
>   /home/lfbrooks/pppa6022/2023/problem_sets/ps1/data/d2010_`date_of_data';

    Result                           # of obs.
    -----------------------------------------
    not matched                             31
        from master                         12  (_merge==1)
        from using                          19  (_merge==2)

    matched                              3,090  (_merge==3)
```

```
    ----------------------------------------

. * fix markers so i can clearly see merging issues *;
. replace d1950 = 0 if d1950 == .;
(19 real changes made)

. replace d2010 = 0 if d2010 == .;
(12 real changes made)

. * look at cross-tab of which counties appear when *;
. tab d1950 d2010;

 1 if year |   1 if year is 2010
   is 1950 |         0          1 |     Total
-----------+----------------------+----------
         0 |         0         19 |        19
         1 |        12      3,090 |     3,102
-----------+----------------------+----------
     Total |        12      3,109 |     3,121

. * list the counties that dont merge *;
. list statefips countyfips year d1950 d2010 name_1950 name if d1950 + d2010 != 2;

      +-------------------------------------------------------------------------+
      | statef~s   county~s   year   d1950   d2010   nam~1950   nam~1950 |
      |-------------------------------------------------------------------------|
1619. |       30        113      .       1       0                       |
1725. |       32        025      .       1       0                       |
2327. |       46        001      .       1       0                       |
2392. |       46        131      .       1       0                       |
2814. |       51        055      .       1       0                       |
      |-------------------------------------------------------------------------|
2848. |       51        123      .       1       0                       |
2851. |       51        129      .       1       0                       |
2862. |       51        151      .       1       0                       |
2881. |       51        189      .       1       0                       |
2891. |       51        560      .       1       0                       |
      |-------------------------------------------------------------------------|
2908. |       51        785      .       1       0                       |
3102. |       56        047      .       1       0                       |
3103. |       04        012   2010       0       1                       |
3104. |       08        014   2010       0       1                       |
3105. |       32        510   2010       0       1                       |
      |-------------------------------------------------------------------------|
3106. |       35        006   2010       0       1                       |
3107. |       51        515   2010       0       1                       |
3108. |       51        550   2010       0       1                       |
3109. |       51        580   2010       0       1                       |
3110. |       51        595   2010       0       1                       |
      |-------------------------------------------------------------------------|
```

```
3111. |       51        600   2010       0       1                       |
3112. |       51        620   2010       0       1                       |
3113. |       51        640   2010       0       1                       |
3114. |       51        678   2010       0       1                       |
3115. |       51        683   2010       0       1                       |
      |-----------------------------------------------------------------|
3116. |       51        685   2010       0       1                       |
3117. |       51        720   2010       0       1                       |
3118. |       51        735   2010       0       1                       |
3119. |       51        775   2010       0       1                       |
3120. |       51        810   2010       0       1                       |
      |-----------------------------------------------------------------|
3121. |       55        078   2010       0       1                       |
      +-----------------------------------------------------------------+

. ****** 4. regression
**********************************************************************;
. * 3(a) *;
. *** bring in data from 1 that I saved ***;
. drop _all;

. use  /groups/brooksgrp/junk/uselater;

. * 3(b) *;
. * population as a function of our four variables, with and without state fixed
effects *;
. regress ln_cv1 s1 s2 s3 s4;

      Source |       SS           df       MS      Number of obs   =     6,211
-------------+----------------------------------   F(4, 6206)      =    464.29
       Model |  2505.48738         4  626.371845   Prob > F        =    0.0000
    Residual |  8372.51911     6,206  1.34910073   R-squared       =    0.2303
-------------+----------------------------------   Adj R-squared   =    0.2298
       Total |  10878.0065     6,210  1.75169187   Root MSE        =    1.1615

------------------------------------------------------------------------------
      ln_cv1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          s1 |  -1.116101   .2210633    -5.05   0.000    -1.549461   -.6827401
          s2 |   .2454641   .2359306     1.04   0.298    -.2170415    .7079697
          s3 |  -10.61731   .3288104   -32.29   0.000     -11.2619   -9.972731
          s4 |    12.6286   .3421147    36.91   0.000     11.95793    13.29926
       _cons |    10.8182   .2242445    48.24   0.000     10.37861     11.2578
------------------------------------------------------------------------------

. regress ln_cv1 s1 s2 s3 s4 i.year;

      Source |       SS           df       MS      Number of obs   =     6,211
-------------+----------------------------------   F(5, 6205)      =    386.01
       Model |  2580.83455         5   516.16691   Prob > F        =    0.0000
```

```
    Residual |  8297.17193     6,205  1.33717517   R-squared       =    0.2373
-------------+----------------------------------   Adj R-squared   =    0.2366
       Total |  10878.0065     6,210  1.75169187   Root MSE        =    1.1564

------------------------------------------------------------------------------
      ln_cv1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          s1 |  -1.172123   .2202106    -5.32   0.000    -1.603812    -.740434
          s2 |   .3030146   .2350106     1.29   0.197    -.1576875    .7637167
          s3 |  -8.913226   .3983665   -22.37   0.000    -9.694163    -8.13229
          s4 |   12.21275   .3450752    35.39   0.000     11.53628    12.88921
             |
        year |
       2010  |   -.605431   .0806539    -7.51   0.000    -.7635406   -.4473214
             |
       _cons |   10.74145   .2234852    48.06   0.000     10.30334    11.17956
------------------------------------------------------------------------------

. * 3(c) *;
. * answer in words *;
. * 3(d) *;
. xi: regress ln_cv1 s1 s2 s3 s4 i.statefips;
i.statefips       _Istatefips_1-49    (_Istatefips_1 for statefips==01 omitted)

      Source |       SS           df       MS      Number of obs   =     6,211
-------------+----------------------------------   F(52, 6158)     =     93.84
       Model |  4809.09983        52  92.4826891   Prob > F        =    0.0000
    Residual |  6068.90666     6,158  .985532097   R-squared       =    0.4421
-------------+----------------------------------   Adj R-squared   =    0.4374
       Total |  10878.0065     6,210  1.75169187   Root MSE        =    .99274

------------------------------------------------------------------------------
       ln_cv1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
           s1 |  -1.612403   .2085785    -7.73   0.000     -2.02129   -1.203516
           s2 |  -.4506948   .2433753    -1.85   0.064    -.9277954    .0264058
           s3 |  -8.476128    .297724   -28.47   0.000    -9.059772   -7.892485
           s4 |   10.38369   .3094779    33.55   0.000      9.77701    10.99038
 _Istatefips_2 |   .2395945   .2096143     1.14   0.253    -.1713227    .6505117
 _Istatefips_3 |   -.341987   .1191919    -2.87   0.004    -.5756447   -.1083293
 _Istatefips_4 |   .7549761    .132003     5.72   0.000     .4962042    1.013748
 _Istatefips_5 |   -.956634   .1285559    -7.44   0.000    -1.208648   -.7046196
 _Istatefips_6 |   1.661874   .2647189     6.28   0.000     1.142933    2.180816
 _Istatefips_7 |   1.078478   .4146538     2.60   0.009      .265612    1.891345
 _Istatefips_8 |   1.766339   .7078582     2.50   0.013       .37869    3.153989
 _Istatefips_9 |   .1072419    .122085     0.88   0.380    -.1320872    .3465711
_Istatefips_10 |  -.6087009   .1022703    -5.95   0.000    -.8091864   -.4082154
_Istatefips_11 |  -1.111877   .1413167    -7.87   0.000    -1.388907   -.8348467
_Istatefips_12 |     -.0232   .1153071    -0.20   0.841    -.2492422    .2028422
_Istatefips_13 |   .0033265   .1181734     0.03   0.978    -.2283347    .2349876
```

```
_Istatefips_14 |  -.2464388   .1172589    -2.10   0.036    -.4763071   -.0165705
_Istatefips_15 |   -1.01042   .1156064    -8.74   0.000    -1.237049   -.7837915
_Istatefips_16 |  -.1639021   .1121195    -1.46   0.144    -.3836954    .0558913
_Istatefips_17 |  -.0388832   .1227779    -0.32   0.751    -.2795708    .2018044
_Istatefips_18 |   .6963108   .1989496     3.50   0.000     .3063001    1.086322
_Istatefips_19 |    .557126   .1677198     3.32   0.001     .2283366    .8859153
_Istatefips_20 |   1.509039   .2090897     7.22   0.000     1.099151    1.918928
_Istatefips_21 |   .1461311   .1203122     1.21   0.225    -.0897228     .381985
_Istatefips_22 |  -.1091943   .1202632    -0.91   0.364    -.3449521    .1265636
_Istatefips_23 |  -.5141518   .1164799    -4.41   0.000    -.7424931   -.2858106
_Istatefips_24 |  -.2831452   .1131821    -2.50   0.012    -.5050216   -.0612689
_Istatefips_25 |  -1.311575   .1330621    -9.86   0.000    -1.572423   -1.050727
_Istatefips_26 |  -1.289386    .118749   -10.86   0.000    -1.522175   -1.056596
_Istatefips_27 |  -1.412422    .194399    -7.27   0.000    -1.793512   -1.031332
_Istatefips_28 |   .7105157   .2409119     2.95   0.003     .2382441    1.182787
_Istatefips_29 |    1.35665     .17826     7.61   0.000     1.007198    1.706102
_Istatefips_30 |   -.727946   .1563599    -4.66   0.000    -1.034466   -.4214259
_Istatefips_31 |   1.089503   .1280204     8.51   0.000     .8385379    1.340467
_Istatefips_32 |    .230987   .1114315     2.07   0.038     .0125423    .4494317
_Istatefips_33 |  -1.268355   .1348613    -9.40   0.000     -1.53273    -1.00398
_Istatefips_34 |   .6013818   .1186587     5.07   0.000     .3687693    .8339943
_Istatefips_35 |  -.5275815   .1231592    -4.28   0.000    -.7690165   -.2861465
_Istatefips_36 |  -.2262173   .1498855    -1.51   0.131    -.5200452    .0676105
_Istatefips_37 |   .9065174   .1259987     7.19   0.000     .6595159    1.153519
_Istatefips_38 |   .7997704   .3274598     2.44   0.015     .1578348    1.441706
_Istatefips_39 |   .1328684   .1349881     0.98   0.325    -.1317554    .3974921
_Istatefips_40 |  -1.307799   .1293623   -10.11   0.000    -1.561394   -1.054204
_Istatefips_41 |  -.0312807   .1153096    -0.27   0.786    -.2573277    .1947664
_Istatefips_42 |   -.744918   .1003851    -7.42   0.000    -.9417079    -.548128
_Istatefips_43 |  -1.121234   .1606454    -6.98   0.000    -1.436155   -.8063124
_Istatefips_44 |   .1941165   .2100361     0.92   0.355    -.2176277    .6058607
_Istatefips_45 |  -.4409798   .1061472    -4.15   0.000    -.6490654   -.2328942
_Istatefips_46 |  -.0176285   .1464438    -0.12   0.904    -.3047095    .2694525
_Istatefips_47 |   .0094313   .1323346     0.07   0.943    -.2499907    .2688534
_Istatefips_48 |   .2541545   .1249985     2.03   0.042     .0091137    .4991953
_Istatefips_49 |   -.763836   .1728659    -4.42   0.000    -1.102714   -.4249584
         _cons |   11.53488   .2333342    49.44   0.000     11.07746     11.9923
------------------------------------------------------------------------------

. xi: regress ln_cv1 s1 s2 s3 s4 i.statefips i.year;
i.statefips       _Istatefips_1-49    (_Istatefips_1 for statefips==01 omitted)
i.year            _Iyear_1950-2010    (naturally coded; _Iyear_1950 omitted)

      Source |       SS           df       MS      Number of obs   =     6,211
-------------+----------------------------------   F(53, 6157)     =    103.84
       Model |  5134.19898        53  96.8716789   Prob > F        =    0.0000
    Residual |   5743.8075     6,157  .932890613   R-squared       =    0.4720
-------------+----------------------------------   Adj R-squared   =    0.4674
       Total |  10878.0065     6,210  1.75169187   Root MSE        =    .96586
```