Econometrics for Public Policy II
Spring 2019

Problem Set 1
Due Class 3 (January 30) 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 of

      - 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

      See Table **??** at end.

   b. Find averages of the same variables by year and state
      See Table **??** at end.

2. Matching Data

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

   3,090

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

   12

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

   19

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

   Menominee County, Wisconson (55/078) was created in 1959 (see Wikipedia).

   La Paz County, Arizona (04/012) was established in 1983 (again, see Wikipedia).

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.)

   Any two reasonable explanations accepted. Generally, many 1950 counties get combined to make 2010 counties; you can cite many such cases.

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 and a fixed effect for year = 2010.

   Results are in the log file.

c. Repeat the previous regression with state fixed effects

   Results are in the log file.

d. Interpret one of the share coefficients from the second 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. Which leads us to the next question...

e. Report how much a one standard deviation change in that share impacts population.

The standard deviation of the share white is 0.16, or sixteen percentage points. A 16 percentage point increase in a county's white population share is correlated with a 33 percent decrease ($= (0.16)(-2.06)$) in overall population.

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

State	year	population	log(pop)	share			
						at least college, age 25+	
				white	African Am.	women	men
AL	1950	45698	10.4	0.668	0.331	0.077	0.063
	2010	71339	10.6	0.683	0.284	0.452	0.406
AZ	1950	53542	10.3	0.824	0.022	0.151	0.141
	2010	426134	11.7	0.731	0.019	0.546	0.524
AK	1950	25460	9.9	0.816	0.184	0.068	0.068
	2010	38879	10.1	0.795	0.161	0.442	0.381
CA	1950	182521	10.8	0.951	0.019	0.168	0.157
	2010	642310	12	0.752	0.033	0.613	0.575
CO	1950	21033	9	0.992	0.003	0.167	0.134
	2010	78581	9.8	0.9	0.016	0.647	0.59
CT	1950	250910	12	0.982	0.017	0.125	0.142
	2010	446762	12.7	0.841	0.065	0.621	0.587
DE	1950	106028	11.3	0.838	0.159	0.104	0.123
	2010	299311	12.5	0.721	0.202	0.545	0.523
DC	1950	802178	13.6	0.646	0.35	0.223	0.284
	2010	601723	13.3	0.391	0.514	0.685	0.685
FL	1950	41363	9.8	0.749	0.249	0.105	0.106
	2010	280617	11.5	0.793	0.145	0.515	0.469
GA	1950	21664	9.5	0.658	0.342	0.085	0.066
	2010	60929	10.2	0.669	0.283	0.45	0.397
ID	1950	13378	9.1	0.988	0.001	0.169	0.148
	2010	35627	9.6	0.928	0.003	0.551	0.532
IL	1950	85413	10.3	0.979	0.021	0.104	0.099
	2010	125791	10.4	0.909	0.05	0.541	0.497
IN	1950	42763	10.2	0.988	0.012	0.101	0.105
	2010	70476	10.6	0.94	0.025	0.464	0.437
IA	1950	26475	9.9	0.997	0.003	0.139	0.104
	2010	30771	9.8	0.954	0.011	0.547	0.488
KS	1950	18146	9.3	0.985	0.014	0.148	0.128
	2010	27173	9.2	0.925	0.018	0.581	0.517
KY	1950	24540	9.7	0.949	0.051	0.078	0.066
	2010	36161	10	0.937	0.037	0.434	0.353
LA	1950	41930	10.2	0.645	0.353	0.081	0.072
	2010	70834	10.6	0.647	0.317	0.434	0.363
ME	1950	57111	10.7	0.997	0.001	0.126	0.101
	2010	83023	11	0.959	0.007	0.563	0.502

				share			
						at least college, age 25+	
State	year	population	log(pop)	white	African Am.	women	men
MD	1950	97625	10.7	0.821	0.179	0.115	0.12
	2010	240565	11.7	0.729	0.2	0.586	0.541
MA	1950	335037	11.9	0.982	0.016	0.145	0.158
	2010	467688	12.4	0.851	0.057	0.654	0.609
MI	1950	76768	10.2	0.981	0.015	0.118	0.097
	2010	119080	10.7	0.91	0.039	0.535	0.5
MN	1950	34281	9.9	0.991	0.001	0.14	0.088
	2010	60965	10.1	0.929	0.013	0.581	0.528
MS	1950	26572	10	0.564	0.435	0.086	0.075
	2010	36187	10.1	0.565	0.41	0.477	0.41
MO	1950	34388	9.7	0.975	0.025	0.097	0.083
	2010	52078	10	0.931	0.035	0.463	0.409
MT	1950	10369	8.7	0.968	0.001	0.194	0.122
	2010	17668	8.9	0.891	0.003	0.586	0.528
NE	1950	14253	9	0.994	0.002	0.132	0.094
	2010	19638	8.7	0.952	0.007	0.584	0.524
NV	1950	9417	8.3	0.925	0.012	0.173	0.151
	2010	158856	9.9	0.865	0.02	0.541	0.513
NH	1950	53324	10.7	0.998	0.001	0.145	0.134
	2010	131647	11.5	0.954	0.008	0.615	0.565
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
NM	1950	21287	9.6	0.931	0.008	0.14	0.133
	2010	62399	10.1	0.762	0.014	0.53	0.493
NY	1950	239197	11.4	0.979	0.019	0.135	0.133
	2010	312550	11.7	0.862	0.062	0.562	0.52
NC	1950	40619	10.3	0.733	0.26	0.103	0.079
	2010	95355	10.9	0.722	0.206	0.538	0.464
ND	1950	11691	9.1	0.977	0	0.147	0.091
	2010	12690	8.6	0.909	0.004	0.586	0.53
OH	1950	90303	10.7	0.974	0.026	0.105	0.104
	2010	131097	11.2	0.926	0.041	0.469	0.436
OK	1950	29005	9.9	0.919	0.05	0.123	0.112
	2010	48719	10	0.772	0.035	0.493	0.454
OR	1950	42259	9.9	0.987	0.002	0.167	0.139
	2010	106419	10.6	0.896	0.007	0.59	0.569

				share			
						at least college, age 25+	
State	year	population	log(pop)	white	African Am.	women	men
PA	1950	156687	11.2	0.983	0.016	0.099	0.102
	2010	189588	11.5	0.918	0.044	0.449	0.43
RI	1950	158379	11.3	0.982	0.017	0.125	0.146
	2010	210513	11.9	0.895	0.032	0.636	0.624
SC	1950	46022	10.5	0.546	0.453	0.107	0.085
	2010	100551	11	0.596	0.361	0.486	0.435
SD	1950	9599	8.8	0.916	0.001	0.163	0.097
	2010	12336	8.7	0.827	0.004	0.564	0.487
TN	1950	34650	9.9	0.903	0.097	0.071	0.061
	2010	66801	10.4	0.894	0.072	0.418	0.372
TX	1950	30359	9.5	0.9	0.1	0.125	0.111
	2010	98998	9.9	0.85	0.064	0.478	0.446
UT	1950	23754	9	0.977	0.001	0.174	0.175
	2010	95306	10	0.921	0.005	0.614	0.595
VT	1950	26982	10	0.999	0.001	0.153	0.114
	2010	44696	10.4	0.96	0.007	0.607	0.516
VA	1950	26131	9.8	0.753	0.246	0.125	0.102
	2010	59709	10.3	0.757	0.192	0.532	0.482
WA	1950	60999	10.1	0.979	0.006	0.171	0.148
	2010	172424	11	0.847	0.012	0.615	0.588
WV	1950	36465	10.1	0.966	0.034	0.088	0.081
	2010	33691	10.1	0.957	0.022	0.404	0.344
WI	1950	48374	10.3	0.99	0.002	0.127	0.092
	2010	78986	10.6	0.924	0.015	0.54	0.494
WY	1950	12105	9	0.987	0.004	0.224	0.164
	2010	24505	9.7	0.931	0.004	0.632	0.564