Clarkwest Data Set of Inequality in U.S. States
June 12, 2009

The data set includes multiple state-level measures of household income inequality and income level for the fifty U.S. states. The measures are derived from decennial Census data covering the years 1970 through 2000. Details on the source data and measures are presented below.

Source Data

The measures are based on U.S Census microdata data obtained via the IPUMS project at the University of Minnesota (Ruggles et al., 2004).

Census Samples

1970 : Form 1 State (1%)
1980 : 5% State (A Sample)
1990 : 5% State
2000 : 1% Census PUMS

Unit of Observation

Statewide measures are calculated using individuals as the unit of observation. Thus, the measures reflect distribution and level of household income across individuals.

Relevant Definitions

Income. Total household income, all sources, 1989 dollars. Census income measures are top-coded, with the level of the topcodes varying across years. Topcodes artificially reduce both measures of mean income level and of income inequality. In order to calculate the measures in the dataset, top-coded incomes were adjusted by a multiplier calculated using the assumption that the top tail of incomes conform to a Pareto distribution.

Racial Categories. In addition to the statewide measures for all individuals, the data set contains some measures for race subgroups. Those include, *non-White*, *White*, and *Black*. The non-White category includes all individuals who self-report their racial identity as anything other than White. The category Black is a subset of non-White, including all individuals who report their racial identity of Black or African American. The category White includes all individuals who self-report their racial identity as White or Caucasian.

Note that for the purposes of Census racial classification, Hispanic is not considered a racial category. Hispanics are represented in all categories above, though they compose a larger proportion of the "all other" category (part of the non-White category above) as many Hispanics do not self-identify with any of the Census racial classifications.

The 2000 Census permitted individuals to report more than one category of racial identity. Consistent with historical convention in the U.S., individuals who reported Black/African American as one of their racial identities were classified as Black for the purposes of calculating subgroup measures of income. All individuals who reported a racial identity other than White were included in the non-White category.

Users should be aware of small sample sizes used to calculate Black and non-White measures in states with low numbers of residents from minority groups. No measures are available for Blacks in Vermont in 1970 because of the presence of only one African American in the Census 1% sample in that year.

Codebook

Number of Variables: 29
Number of Observations: 200

Variable Name	Description
state	FIPS Code (number between 1 and 56)
stname	Two-Letter State Postal Abbreviation
year	Year (four digits)
giniall	Gini coefficient (x100), full sample
gininwt	Gini coefficient (x100), non-Whites
giniwht	Gini coefficient (x100), Whites
giniblk	Gini coefficient (x100), Blacks/African Americans
theil	Theil index, full sample
pctincb50	Percent of total income received by bottom half of distribution, full sample
allinc95p	95th percentile of income, full sample
allinc90p	90th percentile of income, full sample
allinc50p	50th percentile of income, full sample
allinc10p	10th percentile of income, full sample
nwtinc95p	95th percentile of income, non-Whites
nwtinc90p	90th percentile of income, non-Whites
nwtinc50p	50th percentile of income, non-Whites
nwtinc10p	10th percentile of income, non-Whites
whtinc95p	95th percentile of income, Whites
whtinc90p	90th percentile of income, Whites
whtinc50p	50th percentile of income, Whites
whtinc10p	10th percentile of income, Whites
blkinc95p	95th percentile of income, Blacks
blkinc90p	90th percentile of income, Blacks
blkinc50p	50th percentile of income, Blacks
blkinc10p	10th percentile of income, Blacks
ahincall	Mean income, full sample
ahincnwt	Mean income, non-Whites
ahincwht	Mean income, Whites
ahincblk	Mean income, Blacks

File Format

The file (clarkwestineq.csv) is provided as a rectangular comma-delimited text file.

Citing the Data

Users of the data should cite Clarkwest, Andrew. (2008). "Neo-materialist theory and the temporal relationship between income inequality and longevity change." *Social Science & Medicine*, 66: 1871-1881.

Contact

Please contact Andrew Clarkwest (aclarkwest@mathematica-mpr.com) with any questions regarding the data set.