

Earnings Disparity Data: Summary of Methodology

OBJECTIVE

The primary objective of this analysis is to measure earnings disparities by sex, race, and ethnicity at the national and state levels. By “disparity” we mean an unequal distribution of earnings across the various participants in an economy by sex, race, or ethnicity.¹ To ascertain where disparities exist, we used five-year estimates from the American Community Survey (ACS). This allowed us to increase the statistical reliability of the data for less populated states (e.g., Wyoming) and small population subgroups (e.g., Asians).

DATA SOURCE

The data source for this project was the ACS five-year dataset for the 2009-2013 period.² To obtain micro-level data, we accessed Public Use Microdata Sample (PUMS) data³ through the Census Bureau’s American FactFinder website.⁴

The ACS supplements the information in the decennial census with annually updated estimates of the nation’s demographic, economic, housing, and social characteristics. The detailed data that were previously collected from the decennial census “long form” are now produced from the ACS. In contrast to the decennial census, which is a point-in-time count of the population, the ACS describes the average characteristics of the population over a specified period of time. The ACS collects information continuously nearly every day of the year and then aggregates the results over one year, three years, and five years. Advantages of using the ACS for this analysis include:

- *Largest survey:* The ACS is the largest survey in the United States and the only source of small-area statistics on a wide range of demographic, economic, housing, and social characteristics for all communities in the country. Each month a sample of about 290,000 addresses receives a questionnaire. About 3.5 million addresses are surveyed each year.

¹ Inequality does not necessarily indicate discrimination, as other, nondiscriminatory factors may contribute to this imbalance.

² The 2009-2013 dataset was the most current data available when this analysis was conducted in 2015.

³ The PUMS files are a set of untabulated records about individual people or housing units. The Census Bureau produces the PUMS files so that data users can create custom tables. The confidentiality of ACS respondents is protected through a variety of means, ensuring that it is impossible to identify individuals who provide a response, yet making sure the results are still useful. For more information about the confidentiality of the PUMS files, see <https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html>.

⁴ U.S. Department of Commerce, Census Bureau, <http://www.census.gov/programs-surveys/acs/data/pums.html>.

- *High survey participation rate:* Response to the ACS is mandatory. The combination of mail, telephone, and personal visits has produced annual survey response rates as high as 98 percent.⁵
- *Five-year estimates:* The ACS offers “period” estimates that represent data collected over one year, three years, and five years. The primary advantage of using multiyear estimates is the increased statistical reliability of the data for less populated areas and small population subgroups. Accordingly, we chose to use the five-year estimates, which have the greatest statistical reliability.
- *Availability of data for small geographic areas:* The ACS provides a broad set of statistics for many geographical areas, including large areas such as states and congressional districts, and small areas such as census blocks and block groups. Future phases of this project may include metropolitan statistical area (MSA) level data.
- *Availability of industry and occupation data:* The ACS contains data on respondents’ occupations and industries of employment. Future phases of this project may reflect these data.

METHODOLOGY

Our methodology was a two-step process. First, we developed a SAS program to read the ACS microdata and aggregate hundreds of Standard Occupational Classification (SOC) codes into the nine occupation categories used in the Equal Employment Opportunity Employer Information Report (EEO-1) at the national and state levels.⁶ In the SAS program, we converted the SOC codes to the nine EEO-1 job categories using the conversion table produced by the U.S. Census Bureau.⁷ The SAS program then generated an output data file that contained the number of workers and their average monthly earnings⁸ by sex, race, and ethnicity for each of the 20 two-digit North American Industry Classification System (NAICS) industry sectors⁹ by the nine EEO-1 occupations.

⁵ U.S. Department of Commerce, Census Bureau, <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates>.

⁶ U.S. Equal Employment Opportunity Commission, <http://www.eeoc.gov/eeoc/statistics/employment/jobpat-eeo1/glossary.cfm>.

⁷ U.S. Department of Commerce, Census Bureau, Equal Employment Opportunity (EEO) Tabulation, <https://www.census.gov/people/eeotabulation/data/eeotables20062010.html>.

⁸ We calculated monthly earnings based on annual earnings and the number of weeks worked. While the median is commonly used to measure earnings disparities, the mean was used for this analysis. As such, it should be understood that our estimates reflect disparities at about the 60th percentile of the income distribution, not at the 50th percentile (i.e., median).

⁹ See U.S. Department of Labor, Bureau of Labor Statistics, Table 2, NAICS Sectors, <http://www.bls.gov/ces/cesnaics.htm>.

Second, we developed an Excel model to estimate the earnings disparities by sex, race, and ethnicity for each of the 20 two-digit NAICS industry sectors by the nine EEO-1 occupations. With data from the SAS output file, the Excel model calculated the weighted mean monthly earnings for each sex, racial group, and ethnic group by multiplying the number of female, male, Asian, Black, Hispanic, and White workers in each industry and occupation by their average monthly earnings and then dividing the resultant aggregate earnings by the total number of workers in each sex, racial group, and ethnic group. Finally, the earnings disparities were calculated by dividing the weighted mean monthly earnings for female workers by the weighted mean monthly earnings for male workers, and by dividing the weighted mean monthly earnings for Asian, Black, or Hispanic workers by the weighted mean monthly earnings for White workers.

NATIONAL AND STATE TABLES

The Earnings Disparity Data contains two tables per state, as well as two national tables. Each table consists of five columns. Below is an explanation of the data in each column.

Table: Earnings Disparity Relative to Male Workers

- Sex: Female or male.
- Number of Workers: This number includes all full-time and part-time workers.
- Share of Workers: These are the percentages of all workers who are female and all workers who are male.
- Average Monthly Earnings: Monthly earnings include overtime pay, bonuses, and commissions. The average is the weighted mean.
- Earnings per Dollar of Male Earnings: This is the mean monthly earnings for female workers divided by the mean monthly earnings for male workers.

Table: Earnings Disparity Relative to White Workers

- Race or Ethnicity: For federal reporting purposes, Asian, Black, and White are races, while Hispanic is an ethnicity.¹⁰ For this reason, the ACS asks separate questions for race and ethnicity, which allowed us to produce four non-overlapping groups: Hispanic; non-Hispanic Asian; non-Hispanic Black; and non-Hispanic White.
- Number of Workers: This number includes all full-time and part-time workers.
- Share of Workers: These are the percentages of all workers who are Asian, Black, Hispanic, and White. The percentages do not add to 100 because not all races are

¹⁰ Office of Management and Budget, Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity, https://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/re_app-a-update.pdf.

included in our analysis. The following races are not included due to their limited size: American Indian or Alaska Native; Native Hawaiian or Other Pacific Islander; and two or more races.

- Average Monthly Earnings: Monthly earnings include overtime pay, bonuses, and commissions. The average is the weighted mean.
- Earnings per Dollar of White Worker Earnings: This is the mean monthly earnings for Asian, Black, or Hispanic workers divided by the mean monthly earnings for White workers.

LIMITATIONS AND PLANS FOR FUTURE PHASES

The ACS is a survey not a census, meaning that data are collected from a sample of the population rather than from every member of the population. In principle, many random samples could be drawn from the population and each would be subject to sampling variability.

The Earnings Disparity Data represents the first phase of what is planned as a multi-phase project. Future phases may include earnings disparities by industry and occupation, as well as by MSA level. We are aware that in January 2016, the Census Bureau released the 2010-2014 five-year PUMS file, and in January 2017, the 2011-2015 five-year PUMS file will be released. Therefore, future phases of the Earnings Disparity Data project will be based on updated ACS data.

In addition, we anticipate that the next phase will provide a separate analysis of full-time workers. We included part-time workers in the current analysis to capture as many data points as possible; however, part-time workers are typically excluded from standard wage gap analyses.¹¹ We also anticipate that future phases will measure earnings disparities at both the median and the mean since standard wage gap analyses typically measure the gap at the median because the mean can be skewed by a small number of millionaire outliers.

¹¹ The two most widely-used sources for wage gap estimates at the aggregate level (i.e., national and state levels) are based on data from the Current Population Survey (CPS), which is conducted jointly by the Bureau of Labor Statistics (BLS) and the Census Bureau. The first source is BLS's usual weekly earnings data, which are collected from one-fourth of the monthly CPS sample of approximately 60,000 households. See U.S. Department of Labor, Bureau of Labor Statistics, <https://www.bls.gov/cps/earnings.htm>. The second source is the Census Bureau's Annual Social and Economic Supplement to the CPS, which is principally conducted in March of each year on a sample of approximately 98,000 households. See U.S. Department of Commerce, Census Bureau, <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc.html>. While both sources produce earnings estimates for part-time workers, analysts typically use only the estimates for full-time workers—specifically, BLS's estimates of usual weekly earnings for full-time workers or the Census Bureau's estimates of annual earnings for full-time, year-round workers.