# Graduate Track - Evaluating the Gender-Wage Gap

## Background

The U.S. Government Accountability Office issued an interim report on Women in the Workforce on December 15, 2022. It reported that

“… women earned an estimated 82 cents for every dollar that men earned … Specifically, annual median pay for women was an estimated $11,243 less than for men (an estimated $49,532 for women and $60,775 for men …”

The file named “Sample ACS 2021 for LMU.csv” contains an extract of the detailed 2021 ACS data with selected fields.

## Objective

1. Build a statistical model of pay and estimate the gender-wage gap. The wage variable to model is INCWAGE.
2. Comment on the above quoted statement and your model results.
3. Support or refute the claim that women in the U.S. are unfairly compensated compared to men. Develop additional analyses and data visualizations to support your position.
   1. Note: You will not be assessed based on whether you choose to support or refute the claim but on how well you defend your position.
4. Additional Questions to Consider:
   1. What are the possible omitted variables that are not found in the ACS data?
   2. And how would these variables potentially bias the estimated gender pay gap?
   3. How might the gender wage-gap vary across different subpopulations in the data provided?

**You will present your findings along with your position on whether women in the U.S. are unfairly compensated compared to men.**

## Hints

* Think of the analytical models that you want to use and the justification for its use.
* The ACS survey is a representative sample of the entire country. It includes full-time, part-time workers, the unemployed, and people not in the civilian labor force. Military should be excluded.
* You have also been provided with two data dictionaries (OCCSOC Codes.xlsx & IND Codes.xlsx) which can be used to attach industry and occupation information to the ACS Data, use these as you see fit.
* The ACS sample is derived from a complex multi-stage sampling plan drawn from the US population of housing units. Please apply the PWGTP (person’s weight) to adjust for their individual sampling weights.
* The distribution of earnings (INCWAGE) is highly skewed. Use a suitable transformation.
* Many variables are highly correlated. Evaluate if any of these will affect your estimate of the gender gap.
* Avoid over-fitting.
* Consult the ACS data dictionary as needed
* Consult the GAO report as needed

## About the ACS

“The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people. Information from the survey generates data that help inform how trillions of dollars in federal funds are distributed each year.

Through the ACS, we know more about jobs and occupations, educational attainment, veterans, whether people own or rent their homes, and [other topics](https://www.census.gov/programs-surveys/acs/guidance/subjects.html). Public officials, planners, and entrepreneurs use this information to assess the past and plan the future. When you respond to the ACS, you are doing your part to help your community plan for hospitals and schools, support school lunch programs, improve emergency services, build bridges, and inform businesses looking to add jobs and expand to new markets, and more.”

<https://www.census.gov/programs-surveys/acs/about.html>

## Selected Fields in the 2021 ACS

|  |  |
| --- | --- |
| **2021 ACS Person Record** | |
| **Variable** | **Description** |
| REGION | Census region and division |
| STATEFIP | State code based on FIPS coding scheme |
| ADJUST | Adjustment factor for income and earnings dollar amounts (6 implied decimal places) |
| PERWT | Person's weight |
| AGE | Age |
| CITIZEN | Citizenship status |
| YRNATUR | Year of naturalization |
| CLASSWKR | Class of worker (i.e. government, employee, private organization, self-employed) |
| SPEAKENG | Ability to speak English |
| TRANWORK | Means of transportation to work |
| LANGUAGE | Language spoken at home |
| MARST | Marital status |
| MARRNO | Number of times married |
| EDUC | Educational attainment |
| SEX | Sex |
| INCWAGE | Wages and salary income from past 12 months |
| UHRSWORK | Usual hours worked per week during past 12 months |
| WKSWORK1 | Weeks worked during past 12 months |
| ANCESTR1 | Ancestry, first response |
| INCOTHER | All other income from past 12 months |
| INCTOT | Total person's income from past 12 months |
| INCEARN | Total person's earnings from past 12 months |
| INCINVST | Interest, dividend, and rental income from past 12 months |
| EMPSTAT | Employment status |
| DEGFIELD | Field of degree, first entry |
| DEGFIELD2 | Field of degree, second entry |
| INDNAICS | Industry according to North American Industrial Classification System |
| RACE | Race |
| OCCSOC | Occupation, classified according to 2018 Standard Occupational Classification (SOC) |
| TRANTIME | Travel time to work |
| IND | Industry |
| ARRIVES | Time of arrival at work |
| DEPARTS | Time of departure for work |
| NCHILD | Number of own children in the household |
| BPL | Place of birth |