

Studying Social Inequality with Data Science

INFO 3370 / 5371
Spring 2023

Sampling for Population Inference

Learning goals for today

By the end of class, you will be able to

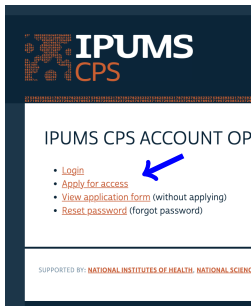
- ▶ explain sampling from a population
- ▶ access survey data online

How to access IPUMS-CPS

1) Visit <https://cps.ipums.org/cps/>. Click **Register**



2) Click **Apply for access**



3) Complete the form

The image shows a registration form titled "NEW IPUMS CPS REGISTRATION". It includes a disclaimer: "IPUMS CPS data are available free of charge. Before using the data, researchers must complete this registration and agree to abide by the usage license specified below. By completing this application, you agree to receive occasional email messages. Such messages will be infrequent, and we will safeguard the confidentiality of your email address." The form fields are: EMAIL (required), PASSWORD (required), PASSWORD CONFIRMATION (required), FIRST NAME (GIVEN NAME) (required), LAST NAME (SURNAME OR FAMILY NAME) (required), and NAME OF INSTITUTION OR EMPLOYER (required) with a dropdown menu.

General research statement: I am in a class using these data to study socioeconomic inequality in America.

How would you design a sample to estimate unemployment?

1. Who would you survey?
2. What mode of data collection?
 - ▶ Mail, phone, web, in person, etc.
3. What if people didn't respond?
4. How to get a high-quality sample while minimizing costs?

How the Current Population Survey is sampled



How the Current Population Survey is sampled



Begin with a **sampling frame**: all housing units in the U.S.

How the Current Population Survey is sampled



Begin with a **sampling frame**: all housing units in the U.S.

- ▶ 1,987 Primary Sampling Units (PSUs)
 - ▶ County or contiguous counties within a state

How the Current Population Survey is sampled



Begin with a **sampling frame**: all housing units in the U.S.

- ▶ 1,987 Primary Sampling Units (PSUs)
 - ▶ County or contiguous counties within a state

- ▶ Stratified (grouped) within states
 - ▶ Stratum: Group of PSUs with similar characteristics
 - ▶ One PSU always chosen per stratum
 - ▶ **Why?** Ensure representation across strata

How the Current Population Survey is sampled



Begin with a **sampling frame**: all housing units in the U.S.

- ▶ 1,987 Primary Sampling Units (PSUs)
 - ▶ County or contiguous counties within a state
- ▶ Stratified (grouped) within states
 - ▶ Stratum: Group of PSUs with similar characteristics
 - ▶ One PSU always chosen per stratum
 - ▶ **Why?** Ensure representation across strata
- ▶ Within PSU, sample geographic clusters of housing units
 - ▶ **Why?** Reduce travel costs for field representatives

More than 75,000 households are sampled

Contacting respondents



Contacting respondents



1. Send a letter

Contacting respondents



1. Send a letter
2. Call or visit in person

Contacting respondents



1. Send a letter
2. Call or visit in person
3. Try many times if needed

Contacting respondents



1. Send a letter
2. Call or visit in person
3. Try many times if needed

Learn about the experience for participants [here](#)

Computer-assisted telephone interview

HELLO

◆ Current Population Survey

Hello. This is from the U.S. Census Bureau.

May I please speak to Respondent name?

- 1 This is correct person
- 2 Correct person called to phone
- 3 Person not home now or not available now (incl. temp ill/hosp.)
- 4 Person unknown at this number
- 5 Person no longer lives there (Includes deceased individuals)
- 6 Other outcome OR problem interviewing household.

[census.gov/programs-surveys/cps/technical-documentation/questionnaires.html](https://www.census.gov/programs-surveys/cps/technical-documentation/questionnaires.html)

Computer-assisted telephone interview

LABFOR

I am going to ask a few questions about work-related activities (THE WEEK BEFORE LAST/LAST WEEK). By (the week before last/last week), I mean the week beginning on Sunday, (DATE), and ending on Saturday, (DATE).

1 Continue

[census.gov/programs-surveys/cps/technical-documentation/questionnaires.html](https://www.census.gov/programs-surveys/cps/technical-documentation/questionnaires.html)

Computer-assisted telephone interview

(THE WEEK BEFORE LAST/LAST WEEK), did (name/you) do ANY work for (pay/either pay or profit)?

- 1 Yes
- 2 No
- 3 Retired
- 4 Disabled
- 5 Unable to work

[census.gov/programs-surveys/cps/technical-documentation/questionnaires.html](https://www.census.gov/programs-surveys/cps/technical-documentation/questionnaires.html)

Annual Social and Economic Supplement

- ▶ Extended survey
- ▶ Conducted each March

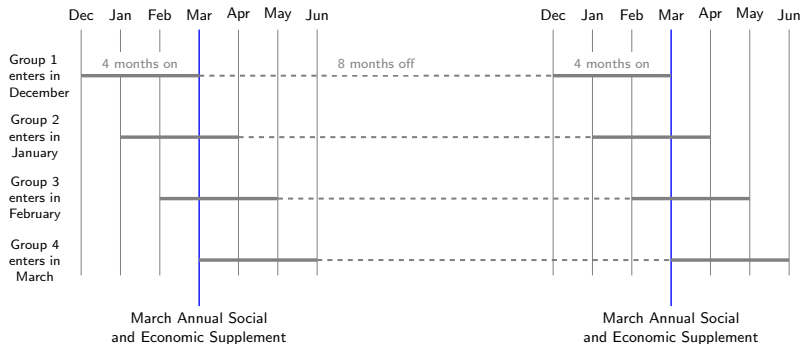
Q48aa

How much did (name/you) earn from this employer before taxes and other deductions during 2021?

- * Enter dollar amount
- * Enter 0 for none

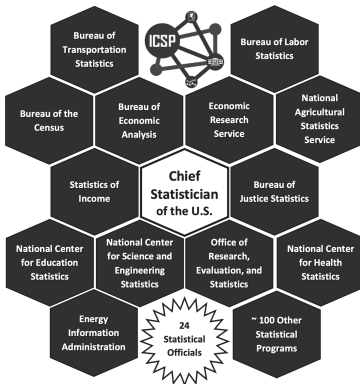
Questionnaire from 2022

Rotating panels



The Current Population Survey (CPS) is only one of many surveys in the federal statistical system

Chart 15-1. THE DECENTRALIZED FEDERAL STATISTICAL SYSTEM



source



The Integrated Public Use Microdata Series (IPUMS) distributes these data and more

- ▶ Easy to access
- ▶ Harmonized documentation
- ▶ Select the variables you want
- ▶ Compare over history



UNIVERSITY OF MINNESOTA

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren and Michael Westberry. Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2022.
<https://doi.org/10.18128/D030.V10.0>



U.S. Census and American Community Survey microdata from 1850 to the present. [Learn More](#)

VISIT SITE



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. [Learn More](#)

VISIT SITE



World's largest collection of census microdata covering over 100 countries, contemporary and historical. [Learn More](#)

VISIT SITE



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#). [Learn More](#)

VISIT SITE



Tabular U.S. Census data and GIS boundary files from 1790 to the present. [Learn More](#)

VISIT SITE



Tabular and GIS data from population, housing, and agricultural censuses around the world. [Learn More](#)

VISIT SITE



Historical and contemporary time use data from 1930 to the present. [Learn More](#)

VISIT SITE



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present). [Learn More](#)

VISIT SITE



Survey data on the science and engineering workforce in the U.S. from 1993 to the present. [Learn More](#)

VISIT SITE

Learning goals for today

By the end of class, you will be able to

- ▶ explain sampling from a population
- ▶ access survey data online