Introduction to Current Population Survey (CPS)

Hsueh-Sheng Wu
Center for Family and Demographic Research
October 18, 2021

BGSU Center for Family and Demographic Research

Outline

- Introduction
- What is special about the CPS?
- Navigate CPS and IPUMS CPS websites
- Survey design
- Subject areas in CPS
  - Basic monthly data
  - Periodic supplemental data
- Analytic tips
  - Using Family Relationship Pointers
  - Using weights
- Studies using CPS data
- Conclusions

Introduction

- The CPS is a monthly survey of about 60,000 households conducted by the U.S. Census Bureau for the Bureau of Labor Statistics since 1962.
- The CPS is the primary source of information on labor force characteristics of the U.S. population, such as employment status, hours of work, and earnings.
- CPS supplemental survey data allow researchers to study topics other than labor force participation and many of these topics interest sociologists.
- Data are collected from households in all 50 states and the District of Columbia. The CPS is representative of the civilian non-institutionalized population.

What Is Special about the CPS?

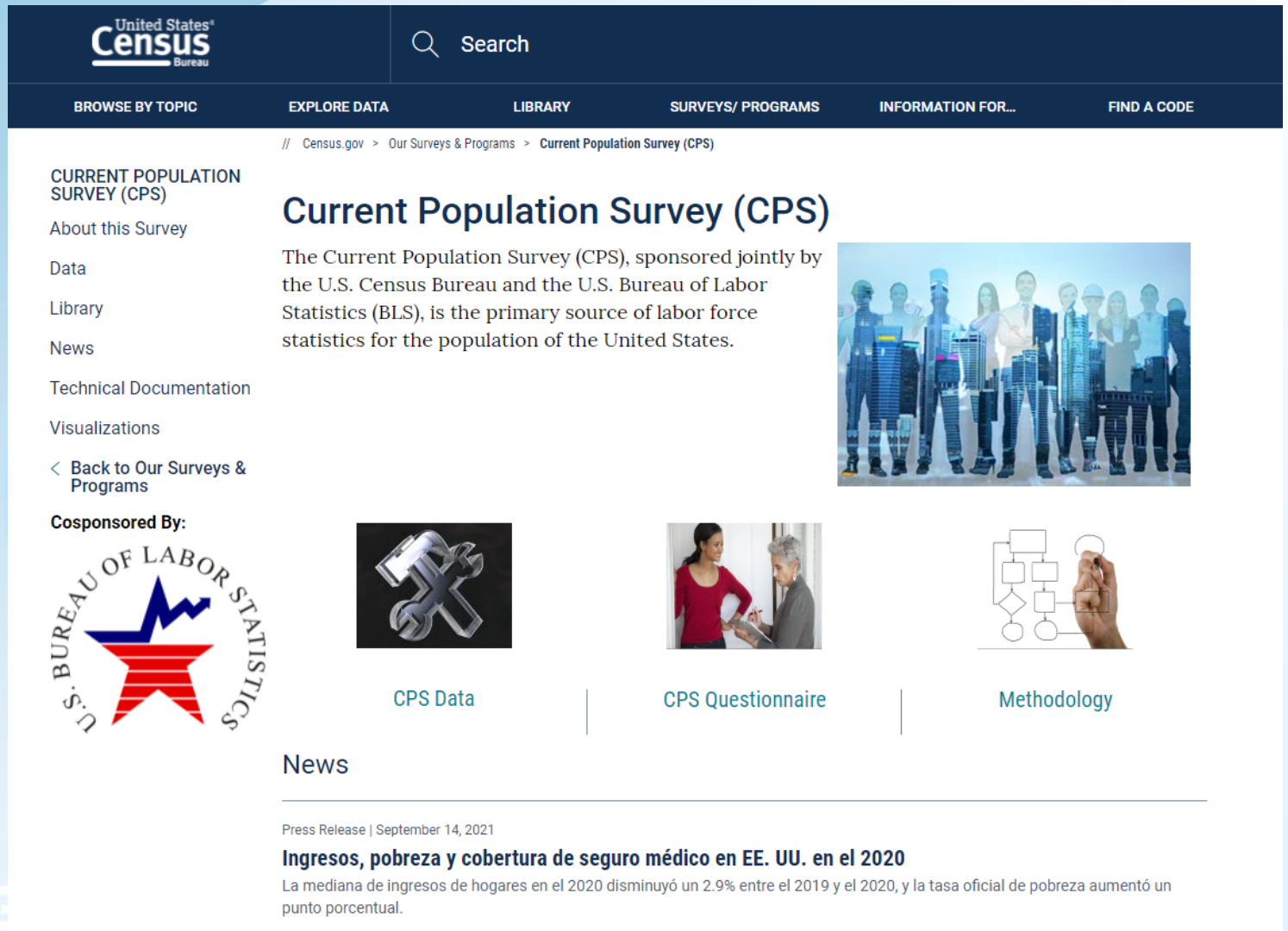
- CPS includes two surveys
  - Monthly surveys collect data on demographic characteristics and labor force
  - Periodic supplemental surveys collect data on many topics such as family characteristics, child support, food security, and others.
- CPS adopted 4-8-4 research design in which respondents are interviewed for four consecutive months and eight months later, re-interviewed for four consecutive months. With this design, the CPS data can be analyzed cross-sectionally or longitudinally.
- Estimates can be obtained at the national, regional, state, or metropolitan (only large metro areas).

What Is Special about the CPS? (Cont.)

- The Annual Social and Economic (ASEC) supplement collects data on relations of family members and allows for better measuring family units.
- Starting in 2007 the CPS provides “pointers” that allow for data line identification of parents and cohabiting partners within a household
- ASEC supplement over-samples or adds the following populations:
  - Armed forces
  - Hispanic sample
  - Children’s Health Insurance Coverage (CHIP) sample
    - Increasing the monthly CPS sample in states with high sampling errors for uninsured children during the February-April using the preceding November CPS sample
- Sociologists have also used CPS supplemental data on food security, disability, fertility, and child support.

Navigate CPS Website

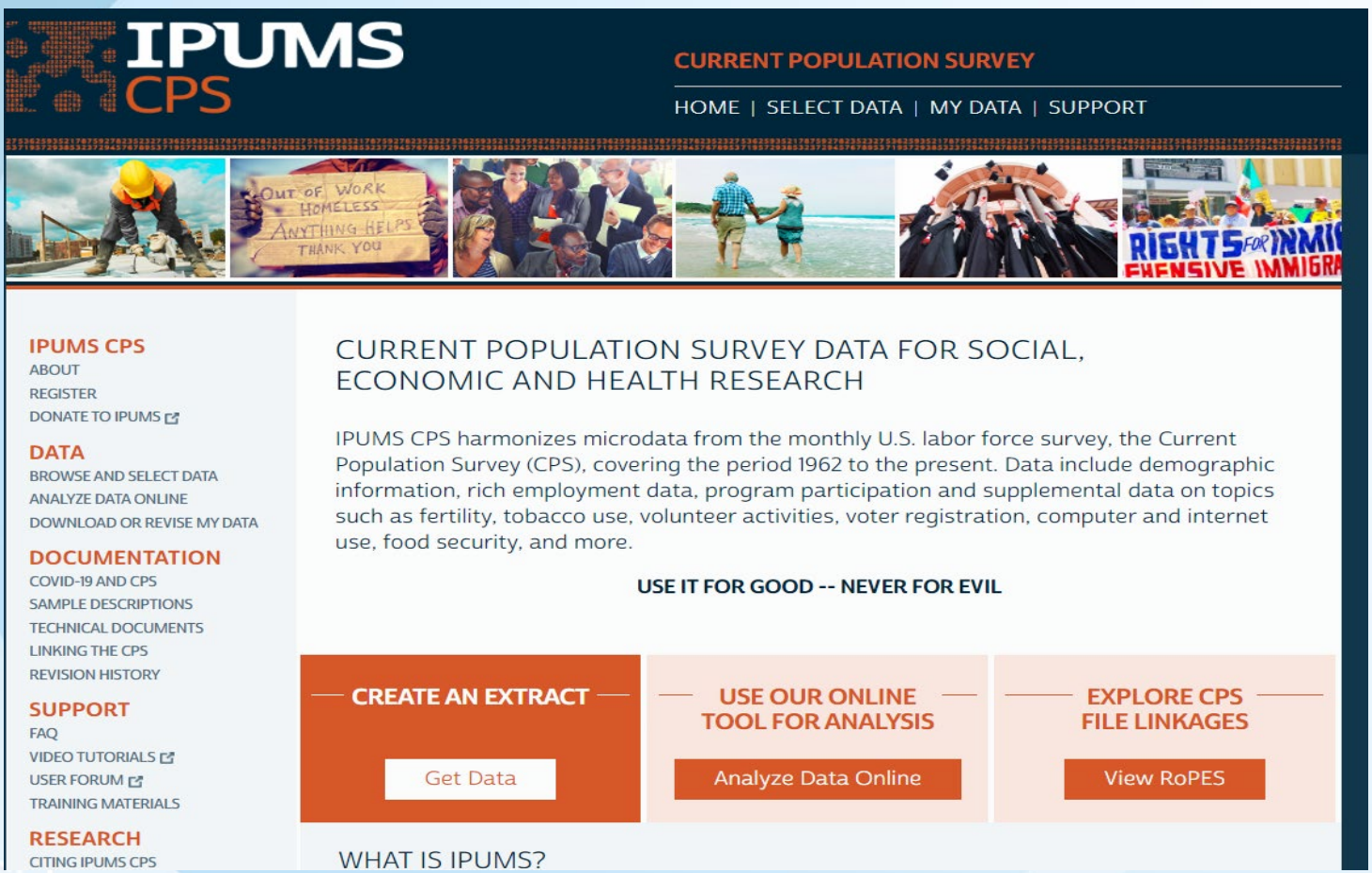
- https://www.census.gov/programs-surveys/cps.html



Center for Family and Demographic Research

Navigate IPUMS CPS Website

- https://cps.ipums.org/cps/



BGSU Center for Family and Demographic Research

Survey Design

- Multi-stage stratified sampling method:
  - 50 states and District of Columbia
    - 792 sampling areas
      - 2,007 counties and independent cities
        - 72,000 housing units or living quarters

- Data are collected with Computer-assisted personal interviewing (CAPI) and questionnaire

Rotation Group Design

Table 1. Rotation Group Design in CPS

	Month in Sample (MIS)							
	1	2	3	4	5	6	7	8
Year 1								
January	A							
Feburay	B	A						
March	C	B	A					
April	D	C	B	A				
May	E	D	C	B				
June	F	E	D	C				
July	G	F	E	D				
August	H	G	F	E				
September	I	H	G	F				
October	J	I	H	G				
November	K	J	I	H				
December	L	K	J	I				
Year 2								
January	M	L	K	J	A			
Feburay	N	M	L	K	B	A		
March	O	N	M	L	C	B	A	
April	P	O	N	M	D	C	B	A
May	Q	P	O	N	E	D	C	B
June	R	Q	P	O	F	E	D	C
July	S	R	Q	P	G	F	E	D
August	T	S	R	Q	H	G	F	E
September	U	T	S	R	I	H	G	F
October	V	U	T	S	J	I	H	G
November	W	V	U	T	K	J	I	H
December	X	W	V	U	L	K	J	I

Rotation Group Design (Cont.)

- Households are interviewed 8 times over 16 months:
  - 4 consecutive months in sample
  - 8 consecutive months out of sample
  - 4 consecutive months in sample

- The households interviewed in the fourth and eight month in sample are referred to as “outgoing rotation groups.” Earnings data are collected from these outgoing rotation groups.

Rotation Group Design (Cont.)

- The 4-8-4 system provides some year-to-year overlap, thus improving estimate of change on both a month-to-month and year-to-year basis; that is, 75% of respondents are the same between successive monthly data and 50% of successive yearly data.

- The rotation group design avoids following respondents for very long time, provides better estimates of change, and avoids discontinuities in the data series

Rotation Group Design (Cont.)

- The rotation group design indicates that the necessity of incorporating the “Month in Sample” variable in creating longitudinal CPS data

Subject Areas in CPS

- Basic Monthly Surveys
  - Employment and unemployment
  - Earnings
  - Hours of work
- Periodic Supplemental Surveys

Table 1. Subject Areas in Periodic supplemental surveys[1]

ASEC	Voter	Displaced Worker	Public Arts
Food Security	Agricultural Worker	Computer and Internet use	Volunteer
Work Schedule	Immunization	Veterans	Civic Engagement
Fertility and Marriage	Child Support	Tabacco Use	Un(der)banked
Education	Job Tenure	Contingent Work	Disability

[1]The availability of Supplemental Surveys can be found on the IPUMS website (https://cps.ipums.org/rotation_pattern_explorer#/)

Analytic Tips

- Using Family Relationship Pointers
- Using Weights

Family Relationship Pointers

- Relations of family members:
- Points of Family Relations
  - Mother
  - Father
  - Spouse and partners
- Identification of all couples in the household
  - Capture far more than the “relationship to head” question
- Previously underrepresented populations:
  - Both Hispanic & both Other couples
  - Both never married couples
  - Couples without children
  - Same sex couples (still small sample)

Using the Pointers

year	month	serial	pernum	relate	age	sex	momloc
2021	3	90749	1	101	55	2	0
2021	3	90749	2	301	12	1	1
2021	3	90750	1	101	52	2	0
2021	3	90750	2	202	53	1	0
2021	3	90750	3	301	18	2	1
2021	3	90750	4	301	16	1	1
2021	3	90752	1	101	34	1	0
2021	3	90752	2	202	34	2	0
2021	3	90752	3	301	10	2	2
2021	3	90752	4	301	7	2	2
2021	3	90752	5	301	2	1	2

year	month	serial	pernum	relate	age	sex	momloc	mom_age
2021	3	90749	1	101	55	2	0	.
2021	3	90749	2	301	12	1	1	55
2021	3	90750	1	101	52	2	0	.
2021	3	90750	2	202	53	1	0	.
2021	3	90750	3	301	18	2	1	52
2021	3	90750	4	301	16	1	1	52
2021	3	90752	1	101	34	1	0	.
2021	3	90752	2	202	34	2	0	.
2021	3	90752	3	301	10	2	2	34
2021	3	90752	4	301	7	2	2	34
2021	3	90752	5	301	2	1	2	34

Stata codes

```
log using "D:\Jason\workshop\CPS 2021\data\pointer.log", replace


import excel "D:\Jason\workshop\CPS 2021\pointer.xlsx", sheet("Sheet1") firstrow clear
list, sepby( year month serial)

* sort the data
sort year month serial pernum

*expand the pointer for mom to the whole family
by year month serial: egen m_momloc = max(momloc)
list, sepby( year month serial)

* Extract the age information for mom
gen t_mom_age = age if pernum==m_momloc
list, sepby( year month serial)

* Expand the age information for mom to the  whole family
by year month serial: egen mom_age = max(t_mom_age)
list, sepby( year month serial)

* keep mom's age only for the children
replace mom_age =. if momloc ~=m_momloc
list, sepby( year month serial)

*remove unncessary variables
drop m_momloc t_mom_age
list, sepby( year month serial)

log close
```

Linking CPS Data

- Why do you need to link the data
  - Examine how these constructs (e.g., employment and food security) are associated with each other
  - Examine how a construct changes over time
  - Examine how these constructs influence each other

Linking CPS Data (Cont.)

- Linking household records is done using household ID variables in CPS.
- Linking records for persons living within households over time is very challenging because CPS does not have a unique longitudinal ID variable for them.
- The National Bureau of Economic Research provides sample command files (http://www.nber.org/data/cps_match.html) to link CPS March data together.
- IPUMS data from Minnesota Population Center (http://cps.ipums.org/cps/) has constructed CPS variables from CPS data sets.
- One advantage of using CPS data from IPUMS is that data from different surveys or different time periods are already linked and users do not need to merge data themselves.

Weighting CPS data

- The reason for weighting CPS data
  - We want accurate mean and standard errors of the estimate.
- Three criteria for choosing weighting variables:
  - Specific weight variables for a CPS data set
  - Types of weight variables: personal weights (i.e., individuals, families, households) or replicate weights
  - Cross-sectional or longitudinal weights
- IPUMS provides an excellent documentations on weighting CPS data
  - https://cps.ipums.org/cps/sample_weights.shtml
  - https://cps.ipums.org/cps/cps_linking_documentation.shtml#linked_weights
  - https://pop.umn.edu/sites/pop.umn.edu/files/final_review_-_cps_stata_exercise_1_0.pdf

Replicate Weight Variables

- Since 2005, Census Bureau releases 160 replicate weight variables for CPS March Supplement data.
- The use of replicate weight variables allows researchers to more accurately estimate the standard error of the parameter estimates.
- The data file that contains replicate weight variables needs to be merged with CPS March Supplement data for analysis
- Special commands are needed for using replicate weight variables in analyzing CPS data.

Stata Command for Weighing CPS Data

- Using ASEC personal weights

```
replace asecwt = round(asecwt)
tabulate foodstmp [fweight=asecwt]
```

Stata Command for Weighing CPS Data (Cont.)

- Using Replicate weights

  svyset [iw=wtsupp], jkrweight(repwtp1-repwtp160, multiplier(.025)) /// vce(jackknife) mse

  svy: tabulate foodstmp

  The -svyset- command describes the survey design of the CPS.

  The -[iw=wtsupp]- command specifies that the sampling weight variable is “wtsupp”.

  The -jkrweight(epwtp1-repwtp160, multiplier(.025))- command instructs Stata that there are 160 replicate weight variables, including repwtp1 through repwtp160 and these variables are used in the Jackknife method to estimate the variance of parameters.

  The -multiplier(.025)- command is decided by the formula provided by Census Bureau.

  The -vce(jackknife) mse - command specifies that a Jackknife method is used to calculate variance and mean square error.

Studies Using CPS data

- The U.S. Census Bureau
  - http://www.census.gov/cps/
- The Bureau of Labor Statistics:
  - http://www.bls.gov/cps/publications.htm
- Integrated Public Use Microdata Series (IPUMS-CPS)
  - http://cps.ipums.org/cps/cpr.shtml

Conclusions

- CPS provides the most recent monthly information on social and economic information in the United States.
- CPS data are collected on the household, family, and individual levels and allow researchers to examine how individuals are influenced by their environments.
- The 4-8-4 design of CPS allows for examining individuals, families, and households both cross-sectionally and longitudinally.
- IPUMS CPS provides more flexibility in downloading and linking CPS data
- CPS and IPUMS CPS websites provide vast information on how to analyze CPS data.
- If you have additional questions, please contact Hsueh-Sheng Wu at wuh@bgsu.edu or 419-372-3119