Overview of the Current Population Survey (CPS)

Sarah Burgoyne
Hsueh-Sheng Wu

Center for Family and Demographic Research

January 23, 2012

BGSU Center for Family and Demographic Research

Outline

- Introduction
- What is special about the CPS?
- Survey design
- Subject areas
  - Basic monthly data
  - Periodic supplemental data
- Data and command files
- Data structure
- Analytic tips
  - Constructing family structure variables
  - Linking CPS
  - 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.

- 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?

- Estimates can be obtained at four geographic levels:
  - National
  - Regional
  - State
  - Metropolitan (only large metro areas)
- Data are collected for each member of the household
- Starting in 2007 the CPS provides “pointers” that allow for data line identification of parents and cohabiting partners within a household

What Is Special about the CPS? (Cont.)

- Annual Social and Economic (ASEC) supplement or CPS March 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 period using the

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

The 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

The 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.

The 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 avoid following respondents for very long time, provides better estimates of change, and avoid discontinuities in the data series.

The 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: Basic Monthly Data

- Employment and unemployment
- Earnings
- Hours of work

Subject Areas: Periodic Supplemental Data

- Annual demographic file
- Employee benefits
- Fertility and immigration
- Health and benefits
- Job training
- Work schedules
- Pensions
- Schooling and computers
- Tobacco use
- Occupational mobility
- Job tenure
- School enrollment
- Veterans
- Voting and registration
- Computer and internet use
- Volunteer workers
- Food security
- Child support
- Lead paint
- Landline and cell phone usage

Obtaining Data Files

- Census DataFerrett
  - http://dataferrett.census.gov/
  - Basic online analyses: one-way frequencies and cross-tabulations
  - Data are also available for download
  - You can download at most 50 variables at one time.
- CPS File Transfer Protocol (FTP)
  - http://www.bls.census.gov/cps_ftp.html
- National Bureau of Economic Research (NBER)
  - http://www.nber.org/data/cps_basic.html
  - http://www.nber.org/data/current-population-survey-data.html
- ICPSR
  - http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/24

Obtaining Command Files

- Census DataFerrett
  - http://dataferrett.census.gov/
- National Bureau of Economic Research (NBER)
  - http://www.nber.org/data/cps_basic.html
  - http://www.nber.org/data/current-population-survey-data.html
- ICPSR
  - http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/24

Data Structure

- CPS data can be conceptualized at three levels: the household level, the family level, and the individual level

Table 2. An Example of Data Structure in the March Supplement data, CPS 2011

Household sequence number (h-seq)	Family type (ftype)	Line number (a-lineno)	Relation to Reference Person (perrp)	Gender (a-sex)	Age (a-age)	Spouse's line number (a-spouse)	Mom's Line Number (pelnmom)	Dad's Line number (pelndad)	type of Mother (pemomtyp)	Type of Dad (pedadtyp)
1	Nonfamily householder	1	Reference person only	Male	55	.	.	.	.	.
2	Primary family	1	Reference Person	Female	31	2	.	.	.	.
2	Primary family	2	Spouse	Male	34	1	.	.	.	.
2	Primary family	3	Child	Male	5	.	1	2	Biological	Biological
2	Primary family	5	Child	Male	13	.	1	2	Biological	Step
2	Primary family	4	Child	Male	10	.	1	2	Step	Biological
3	Primary family	1	Reference Person	Female	55	2	.	.	.	.
3	Primary family	2	Spouse	Male	56	1	.	.	.	.
3	Related subfamily	3	Child	Male	31	4	1	2	Biological	Biological
3	Related subfamily	4	Other rel. of ref. person	Female	31	3	.	.	.	.
3	Related subfamily	5	Grandchild	Female	7	.	4	3	Biological	Biological

Note: The value of 0 in spouse's line number and the value of -1 for line number/type of mother or father were coded as missing to make the table more readable.

Analytic Tips

- Using Family Structure Pointers
- Linking CPS files
- Using Weights

Spouse Pointers

- Identification of unmarried partners:
  - Boyfriend
  - Girlfriend
  - Partner
- Identification of all couples in the household
  - Captures 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)

Parental Pointers

- Number of Parents in a Household
  - Spouse pointer: married versus unmarried parents
- Parents: biological, step, or adoptive
- Siblings: biological, step, half, or adoptive
- CPS & SIPP are the only Census surveys from which it is possible to get estimates of children living with unmarried parents who are cohabiting even if neither parent is the householder
- Geographic level variables allow for comparison tables
  - These data are not suitable for ranking, only general comparisons

Using the Pointers

Original Data

	peridnum	fh_seq	age	sex	pelnmom	pelndad
parents	500101	90813	35	female	-1	-1
	500102	90813	31	male	-1	-1
kids	500103	90813	14	male	1	2
	500104	90813	12	female	1	2
	500105	90813	6	female	1	2

Transformed Data

	peridnum	fh_seq	age	sex	mom_age	dad_age
kids with parent information	500103	90813	14	male	35	31
	500104	90813	12	female	35	31
	500105	90813	6	female	35	31

BGSU Center for Family and Demographic Research

It Gets Complicated…

Original Data

	peridnum	fh_seq	age	sex	pelnmom	pelndad
parents / grandparents	403101	12345	75	female	-1	-1
	403102	12345	76	male	-1	-1
parents & kids	403103	12345	50	female	1	2
	403104	12345	18	female	3	-1
kid	403105	12345	15	male	3	-1
parents & kids	403106	12345	47	female	1	2
	403107	12345	49	male	-1	-1
kids	403108	12345	13	male	6	7
	403109	12345	12	female	6	7
	103110	12345	2	male	4	-1

Transformed Data

Transformed Data

peridnum	fh_seq	age	sex	mom_age	dad_age
403103	12345	50	female	75	76
403104	12345	18	female	50	.
403105	12345	15	male	50	.
403106	12345	47	female	75	76
403108	12345	13	male	47	49
403109	12345	12	female	47	49
103110	12345	2	male	18	.

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
- Three types of linking:
  - Linking CPS data cross-sectionally
  - Linking CPS data longitudinally
  - Linking CPS data both longitudinally and cross-sectionally

Linking CPS Data (Cont.)

- Linking household records is not difficult because CPS provides household ID variables
- Linking records for persons living within households over time is very challenging because CPS does not have an 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.
- Minnesota Population Center (http://cps.ipums.org/cps/) provides web-tools to extract constructed CPS variables from CPS data sets

Weighting CPS data

- The reason for weighting CPS data
  - We want accurate mean and standard errors of the estimate

- Three criteria for choosing the weighting variables:
  - CPS data set
  - unit of analysis
  - whether the data set provides the replicating weight variables

Select Weight Variables in CPS

Table 3. Select weighting variables in CPS core data, March Supplement data, and December Supplement data.

Variable Name	Description	Purpose
Monthly Core data		
HWHHWGT	Household Weight	Used For Tallying Household Characteristics
PWFMWGT	Family Weight	Only Used For Tallying Family Characteristics.
PWSSWGT	Final Weight	Used For Most Tabulations, Controlled To Independent Estimates For 1) States; 2) Origin, Sex, And Age; And 3) Age, Race, and Sex.
PWORWGT	Outgoing Rotation Weight	Used For Tallying Information Collected Only In Outgoing Rotations (i.e., Earnings)
PWVETWGT	Veterans Weight	Used For Tallying Veteran's Data Only
PWCMPWGT	Composited Final Weight.	Used To Create Labor Force Statistics Published by Bureau of Labor Statistics
March Supplement data		
HSUP-WGT	March Supplement Household Weight	Used For Tallying Household Characteristics
FSUP-WGT	March Supplement Family Weight	Only Used For Tallying Family Characteristics.
MARSUPWT	March Supplement Final Weight	Used For Most Tabulations, Controlled To Independent Estimates For 1) States; 2) Origin, Sex, and Age; And 3) Age, Race, And Sex.
A-FNLWGT	Basic Cps Personal Weights (Excluding Spanish Sample)	Used For Most Tabulations, Controlled To Independent Estimates For 1) States; 2) Origin, Sex, And Age; And 3) Age, Race, and Sex.
A-ERNLWT	Basic Cps Earnings Weight	Used For Tallying Information Collected Only In Outgoing Rotations (i.e., Earnings)
December Supplement Data		
HHSUPWGT	Supplement Weight For The Household	Used For Tallying Household Characteristics
PWSUPWGT	Supplement Person Weight for Each Household Member	Used For Most Tabulations, Controlled To Independent Estimates For 1) States; 2) Origin, Sex, And Age; And 3) Age, Race, and Sex.

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 Replicate Weight Variables

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

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
- The National Bureau of Economic Research
  - http://www.nber.org/
- Articles & Databases search through BGSU library website
  - EBSCO Academic Search Complete
  - JSTOR
- Integrated Public Use Microdata Series (IPUMS-CPS)
  - http://cps.ipums.org/cps/cpr.shtml
- The Interuniversity Consortium for Political and Social Research (ICPSR) website
  - http://www.icpsr.umich.edu/icpsrweb/ICPSR/biblio/resources?collection=DATA&q=cps

Conclusions

- CPS helps 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.
- To construct new variables for family structures, researchers need to understand the data structure of the CPS and use those pointer variables
- Researchers can link CPS data together to examine the change in households, families, and individuals.

Conclusions (Cont.)

- CPS data are not collected with simple random sampling methods. When CPS data are analyzed, they need to be weighted with accurate weight variables.

- For further help with using CPS data
  - Find technical documents on the webpage: http://www.census.gov/apsd/techdoc/cps/cps-main.html
  - Contact the Housing and Household Economic Statistics Division at (301) 763-3242 or the Current Population Survey Branch at (301) 763-3806.
  - Visit ask.census.gov for further information on the Current Population Survey/Annual Social and Economic (ASEC) Supplement
  - Contact Hsueh-Sheng Wu at wuh@bgsu.edu or 419-372-3119