CPS Labor Extracts

1979 - 2001

NBER

Apr 2002

Topic	Page
Introduction	2
Miscellaneous variables	7
Geography	10
Demography	14
Wages	24
Employment	27
Union Status	35
Crosswalk table	36

(Appendices are on disk in directory /docs)

CPS Labor Extracts

1979 - 2001

Daniel Feenberg
Jean Roth1

April 2002

http://www.nber.org/data/morg.html

Abstract

The Current Population Survey (CPS) is the government monthly household survey of employment and labor markets. It is the source of the unemployment rate announced each month in the popular press. Since 1968 public use micro data files have been available from the Bureau of Labor Statistics for external analysis. In the interest of ease of use, the NBER has prepared a CD-ROM with extracts of the files from 1979 on.

The extracts include individual data for about 30,000 individuals each month. The 50 or so variables selected relate to employment: hours worked, earnings, industry, occupation, education, and unionization. The extracts also contain many background variables: age, sex, race, ethnicity, geographic location, etc. Annual income is not among the variables - that question is asked only in March. Aside from standardizing the many different codes used by Census to indicate missing values, most variables are just as created by Census. In a few cases (noted in the documentation) variables have been recoded to enhance uniformity through time.

Credits

These extracts were initiated by a collective effort of a number of researchers. Dan Feenberg developed and maintained these extracts for many years. Jean Roth began preparing these extracts in March 2000 and made the code Y2K compliant. Jean Roth and Dan Feenberg are responsible for all errors and this documentation. Special thanks to Inna Shapiro, William Gould, David Autor, Danny Blanchflower, David Macpherson, and Alida Castillo-Freeman. Questions, suggestions, and corrections should be sent to Jean Roth at jroth@nber.org.

---

1 Please report errors or send comments or suggestions to Jean Roth at jroth@nber.org

Sample:

The Current Population Survey (CPS) is a monthly survey of about 60,000 households. An adult (the reference person) at each household is asked to report on the activities of all other persons in the household. There is a record in the file for each adult person. The universe is the adult non-institutional population.

Each household entering the CPS is administered 4 monthly interviews, then ignored for 8 months, then interviewed again for 4 more months. If the occupants of a dwelling unit move, they are not followed, rather the new occupants of the unit are interviewed. Since 1979 only households in months 4 and 8 have been asked their usual weekly earnings/usual weekly hours. These are the outgoing rotation groups, and each year the BLS gathers all these interviews together into a single *Merged Outgoing Rotation Group File*. A consequence of this construction is that an individual appears only once in any file year, but may reappear in the following year.

If you append records from the next year you will get repeated observations on the same individual, and you would want to worry about your standard errors, possibly using the Huber option on the regression command.

The BLS calls these files the *Annual Earnings Files*, but we prefer the name *Merged Outgoing Rotation Groups*, because there is no information in the file on annual earnings. Only hourly or weekly earnings are recorded.

The sample is stratified to provide better estimates for minorities and smaller political jurisdictions. Weights are provided for the preparation of descriptive values and tabulations.

All persons 16 years of age or over are included in the extracts.

The Census Bureau and Bureau of Labor Statistics recently released a major update of CPS Design and Methodology, Technical Paper 63. A pdf copy is available at http://www.bls.census.gov/cps/tp/tp63.htm.

CD-ROM Structure:

The data are provided as a series of annual STATA .dta files. Each file contains all outgoing rotation groups for a single year. From within STATA any file can be loaded with a use statement. For example, if the CD-ROM is drive D:, then the statements:

```
set memory=32m
use d:\morg\annual\morg79
```

will load the entire 1979 file. As each year is 25-28 megabytes, you may wish to restrict the data loaded. Here is an example that retrieves two variables for January only:

```
use weight veteran if intmonth==1 using d:\annual\morg79
```

Value labels are available for most of the variables in the \sources\labels directory. To use the Stata value labels, type 'do d:\sources\labels79_82'. To clear a label such as race, type 'label drop race'. SAS and SPSS value labels are also included in the \sources\labels directory.

Danny Blanchflower has graciously contributed STATA do files which provide statewide unemployment rates and many value labels. You can incorporate this into your working file with: do d:\sources\morg79.

Alternatives to STATA:

As noted, the extracts are Stata binary save files. These files are compact and portable across operating systems and hardware platforms. Non-Stata users can use a conversion program such as STAT/Transfer to translate the Stata files into other formats. For example: copy morg79.dta morg79.tpt will generate a SAS transport file. Conversion software supports many statistical and database packages. Only recent versions of STAT/Transfer will correctly read Stata files generated in non-Intel byte order, such as these.

Complete copies of the entire content of the raw data files are available from http://www.nber.org/data/cps_basic.html or Unicon Inc.

Vendors Mentioned:

Stata Corporation
702 University Drive
College Station TX 77840
409-696-4600
800-782-8272
Stata@Stata.com
http://www.Stata.com

Publications Department
NBER
1050 Mass. Ave.
Cambridge MA 02138
617-868-3900
orders@nber.org
http://www.nber.org

Circle Systems (Stat/Transfer)
1001 Fourth Ave Place #3200
Seattle WA 98154
206-682-3783
stsales@circlesys.com
http://www.stattransfer.com

Unicon Inc.
1640 Fifth Street
Santa Monica CA 90401
310-393-4636
http://www.unicon.com

The data dictionary:

In the dictionary below, for each variable a header line gives:

1. The variable name in the 1989 CPS documentation from the BLS, and below that the name for 1994 on.
2. The variable name in the CD-ROM STATA .dta files.
3. The range of values for that variable.
4. The years for which that variable is available.
5. The universe for non-missing values.

Following the header is a description of the variable, and the possible values it may take on. Sometimes a variable definition changes through time, which will be noted. Major changes in variable definitions have led to the creation of distinct variable name, usually by appending a two-digit year to the variable name. Small changes are tolerated and noted in the description. The source for all variable documentation is from the 1978, 1982, 1984, 1985, 1986, 1989, 1992, 1994, 1995, and 1998 versions of ``Attachment A of the Current Population Survey Interview Record Layout, BLS Microdata File, Basic Monthly Survey, (January.)'' CPS Documentation for March Annual Demographic File is very different. Copies of the CPS layouts are on the CD-ROM in .PDF format, in the ./docs directory.

Miscellaneous variables

h-id	hhid	12 digits	79 - 95:8 all
hrhhid		15 digits	95:9 -

1979 - 1995	Digits 1-2	- regional office number
	Digits 3-5	- PSU
	Digits 6-9	- segment
	Digits 10-12	- household serial number
1995 -	Digits 13-15	- Census county code

Item 9. Household id along with minsamp and intmonth and after 1993, state, is a unique household identifier. Hhid does not have the documented digit structure from 1995:7-1995:9 due to sample resdesigns. It is just a family sequence number (but not sorted).

This survey is structured so that an adult in a dwelling unit is interviewed once a month for four months (minsamp=1-4). Then that dwelling unit is ignored for eight months, and then an adult at that dwelling is interviewed again once a month for four months (minsamp=5-8). If the occupants move, the new occupants are interviewed.

The usual weekly earnings/usual weekly hours are asked only in minsamp=4 and minsamp=8, the last month of each four-month round of interviews. These are the minsamps that are included in this extract. This means that a typical dwelling unit will be included twice, once a year for two years.

Programs on longitudinal matching of CPS respondents by Madrian and Lefgren, http://papers.nber.org/papers/T0247, are available in /docs/matching. Every recent CPS March Annual Demographic File documentation set includes a section on matching CPS samples across years, the main point being that while matching households is supported by the household id, there are no individual identifiers before 1994, so within households matches must be done with age, intmonth, and sex. Finally, there is no matching possible between January to September 1985 and 1986, or between July to December 1984 and 1985, or between June to December 1994 and 1995,or between January to August 1995 and 1996 because of sample redesigns.

a-lineno pulineno	lineno	01-99	79-	all

Item 18a. Person Line Number in household. Supposedly useful in matching individuals across years. Before 1994 when a household member departs other members may change line number.

h-mis hrmis	minsamp	4 or 8	79-	all

Month in Sample. Each household entering the CPS is interviewed for 4 months, then ignored for 8 months, then interviewed again for 4 more months. So for any household minsamp 8 occurs exactly one year after minsamp 4. Only households in interview months 4 and 8 are asked their usual weekly earnings/usual weekly hours. So a typical household appears precisely twice in an outgoing rotation group, and those are the only households included in the extracts.

Hrlonglk	hrlonglk	0,2	94-	all

Longitudinal Link Indicator. A replacement household has no members of the original household living at this address. Note that this variable is not very useful since it refers to a replacement with respect to the prior month, not prior year.

Replacement household	0
Continuing household	2

h-year	year	79-	79-	all

Interview year.

h-month hrmonth	intmonth	01-12	79-	all

Interview calendar month. Matching households in successive years should have the same intmonth. A few do not, reasons unknown.

January	01
...	
December	12

h-hhnum huhhnum	hhnum	1-8	79-	all

Household ID. Matching households should have the same hhnum. This variable notes which household is living at this address. The household interviewed in the first month gets a 1. If a new household moves in, it gets a 2 and so on.

a_fnlwgt pwsswgt	weight	0-20549	79-	all

This is the Final Weight. The sum of the Final Weights in each monthly survey is the US non-institutional population. The CD-ROM excludes persons under 16 years of age. The outgoing rotation group includes one-fourth of that population. So one single month MORG file is one-fourth the population 16 years of age and over, and a year of MORG would sum to 3 times that population. Zero weights appear in some years, for records of unknown function. The implied two or four (1994 on) decimals on the tapes are explicit here.

a-ernlwt pworwgt	earnwt	836-77523	79-	all

Earnings weight for all races. Used for tabulating earnings related items. Since the CD-ROM includes all persons asked earning questions, this sums to the total population each month and 12 times the population for each MORG file. This is not precisely 4 times the weight, presumably because the Census has external knowledge of the size and composition of the labor force. The implied decimals on the tapes are explicit here. A BLS letter suggests that this weight is preferred for all purposes.

Geography

hg-st60 state 11-95 79- all
gestcen

1960 Census Code for state. First digit of state code is division code. These codes do not change.

New England Division
Maine 11
New Hampshire 12
Vermont 13
Massachusetts 14
Rhode Island 15
Connecticut 16

Middle Atlantic Division
New York 21
New Jersey 22
Pennsylvania 23

East North Central Division
Ohio 31
Indiana 32
Illinois 33
Michigan 34
Wisconsin 35

West North Central Division
Minnesota 41
Iowa 42
Missouri 43
North Dakota 44
South Dakota 45
Nebraska 46
Kansas 47

South Atlantic Division
Delaware 51
Maryland 52
D.C. 53
Virginia 54
West Virginia 55
North Carolina 56
South Carolina 57
Georgia 58
Florida 59

East South Central
Kentucky 61
Tennessee 62
Alabama 63
Mississippi 64

West South Central
Arkansas 71
Louisiana 72
Oklahoma 73
Texas 74

Mountain
Montana 81
Idaho 82
Wyoming 83
Colorado 84
New Mexico 85
Arizona 86
Utah 87
Nevada 88

Pacific
Washington 91
Oregon 92
California 93
Alaska 94
Hawaii 95

The city coding system changes in October 1985 from one based on 57 SMSA identifiers with each SMSA divided into a central city and non-central city component to a more complex system of 252 CMSA (Consolidated Metropolitan Statistical Areas) identifiers, some subdivided into as many as 12 PMSAs (Primary Metropolitan Statistical Areas) and up to 5 different Individual Central City Codes. In April of 1994 the rank codes for cities are dropped, but the MSA FIPS codes are retained. In 1995, the 1993 modification to the MSA/FIPS codes are adopted. The BLS has warned that all SMSA coding for 1995 is suspect. Users should understand that the geographic coverage of metropolitan areas increases through time, and not only in Census years. Lists of metropolitan identifiers are on the CD-ROM in /docs. These values are supplied by Census until 1994, when telephone interviews start. After that the respondent is asked their address.

Changes in Metropolitan Areas, 1950-1994, (metrochg.pdf in /docs) lists each metropolitan area in the CPS, the counties that comprise the MAs, and the changes in the MAs' county composition over time. A handful of MAs have been added, or added to, since the writing of the chapter above. 1990 Land Area for Metropolitan Areas (1996 Definition) lists these changes (gead9498.pdf in /docs).

h-metsta gemetsta	smsastat	1-2	79-	all

Metropolitan Status Code. The status of any given location may change in 1986. Not identified was coded as 3 or -1 on the BLS tapes.

Metropolitan	1
Non-metropolitan	2
Not identified	missing

hg-msas gemsast gtmsast	centcity	1-3	79-95:5, 95:9-	all

Central City Code. This looks like more information than smsastat, but many records identified in smsastat are not identified here. Not Identifiable was coded as 4 or -1 on the BLS tapes. This code is missing June, July, and August of 1995.

Central City	1
Balance of SMSA	2
Non SMSA	3
Not identifiable	missing

na	smsa70	1-2	79-85:6	SMSAs

1970 Census SMSA size categories. From April 1984 to July 1985, a new CPS design was phased in. See cpsmar85.pdf at www.nber.org/data/cps.html for more detail. See next entry for same variable after September 1985. This code is missing July to September 1985.

3 million plus	1
1-3 million	2
Not identifiable	missing

hg-mssz	smsa80	2-8	85:10-94:3	SMSAs
gemsasz gecmsasz	smsa93	2-7	95:9-	SMSAs

Reflects 1983 population estimates for the MSA/CMSA. In the original tape, 0 and 1 are used for missing values before 1994, then -1. In 1994 this becomes the population of the CMSA/MSA and the 2 largest categories are combined. This code is missing for April 1994 to August 1995. See /docs/usernote.asc for more detail.

Not identified	missing
100,000-249,000	2
250,000-499,999	3
500,000-999,999	4
1-2.5 million	5
2.5-5 million	6
5 - 10 million	7
10 million plus	8

na	smsarank	0-57	79-85:6	all

The CPS uses the 1970 Census ranking to identify SMSAs from 1973 to 1985. See d:\sources\labelsYY.do or Appendix E for codes. This value is missing for all records during the 3rd quarter of 1985, and the cmsarank variable starts in the 4th quarter - no similar information is provided for 1985:7-9.

Not an SMSA	0
1970 rank	1 - 57

hg-msar gemsark	cmsarank	1-252	86-94:3

CMSA/MSA Rank Code. See Appendix F List 1 for list of codes

Not an CMSA	missing
1980 rank	1 - 252

hg-pmsa gepmsrk	pmsarank	1-12	86-94:3	

PMSA rank code identifies PMSAs within a CMSA. See Appendix F List 2 for codes.

non-divided CMSA PMSA code		missing 1 - 12

h-inducc geindvcc	icntcity	1-4	86-	

Individual Central City Codes identify individual central cities within CMSAs with more than one central city. See Appendix F List 3 for codes.

Other 1980 CC code		missing 1 - 4

hg-msac gemsa gtmsa	msafips	80-9340 80-9360	89-94 95-95:5, 95:9-	smsastat=1

Metropolitan Statistical Area FIPS code. See labelsYY.do or Appendix F List 4 for codes. This code is missing for June, July, and August of 1995.

Not an MSA or not identified	0
1980 CC code	80 – 9340 or 80-9360

hg-cmsa gecmsa gtcmsa	cmsacode	7-91 7-97	89-93 94:1-94:3, 95:9-	

Consolidated Metropolitan Statistical Area Code. See labelsYY.do or List 5 of Appendix F. This code is missing April 1994 to August 1995. See /docs/usernote.asc for more detail.

not a CMSA 1980 CMSA code		0 7-91 or 7-97

Demography

a-sex pesex	sex	1-2	79-	all

Item 18g for 84-88. There are missing values in 1985, and 1989 on.

male	1
female	2

na	race	1-3	79-88	all
a-race	race	1-5	89-95	all
perace		1-4	96-	

'What is ... race?' More race detail is offered for 1989 on. There is no 'other' category for 1996 on, because the Census Bureau began to allocate all 'other' responses into one of the 4 main race categories.

Item 18J.

	79-88	89-95	96-
White	1	1	1
Black	2	2	2
American Indian		3	3
Asian or Pacific Islander		4	4
Other	3	5	

a-reorgn prorigin	ethnic	1-9	79-	all

Item 18k. ‘What is the origin or descent of ...?’ This variable subdivides the Hispanic community by national origin of ancestry. Non-Hispanics were sometimes coded as `A' or '10' on the original BLS tapes. In the extracts non-Hispanic is coded always as '8'. In 1994 only undocumented values of 11-13 appear.

Mexican American	1
Chicano	2
Mexicano	3
Puerto Rican	4
Cuban	5
Central or South American	6
Other Spanish	7
All other	8
Don’t know	9

a-age peage	age	16-99	79-	all

Years of age. The CPS documentation claims that this is topcoded at 90 years of age, but values up to 99 are found for 1979-1985. For 1994 on, this is derived from a question about date of birth.

a-maritl marital 1-7 79- all
prmarsta

Item 18e. Marital status at time of enumeration. Until 1989 Widowed Divorced and separated were grouped, however in all years, <4 is married, otherwise single. In the original data 5 is used for Never Married until 1989.

Married civilian spouse present	1
Married AF spouse present	2
Married spouse absent or separated	3
Widowed or divorced(Through 88)	4
Widowed (After 88)	4
Divorced "	5
Separated "	6
Never Married	7

a-pfnocd ownchild 0-8 84-93, 99:11- all
prnmchld

Number of own children less than 18 in primary family. We code the actual number of children. In the original files, "Not in primary family" is zero and a-pfnocd is the number of own children plus one. Note that all members of the primary family (parents, children, and others alike) get the same value for number of own children in the primary family. Use pfamrel or relref to identify family relationships.

Not in primary family	.
Number of children	0-8

a-pfprcd	chldpres	0-8	84-88	all
prchld		0-15 or 16	89-93, 99:11-	

Presence of own children less than 18 in primary family. As with number of own children, all primary family members get the same value for this variable. Use pfamrel or relref to identify family relationships. This variable is in effect the convolution of 3 or 5 dummy variables into a single two-digit field. In the extracts we have deconstructed this variable into the 5 underlying variables (see below).

	84-88	89-93	99:11-
Not in primary family	0	0	
No children < 18 years old	1	1	0
All children 14-17 years old	2	5	4
All children 6-13 years old	3	4	3
All children 0-5 years old	4	6	5
Children 6-17 (no 0-5)	5	11	10
Children 0-5 and 14-17 (no 6-13)	6	13	12
Children 0-13 (no 14-17)	7	12	11
Children from all age groups	8	16	15
All children 0-2 years old		2	1
All children 3-5 years old		3	2
Children 0-2 and 6-13 (no 3-5 or 14-17)		7	6
Children 0-2 and 14-17 (no 3-13)		8	7
Children 3-5 and 6-13 (no 0-2 or 14-17)		9	8
Children 3-5 and 14-17 (no 0-2 or 6-13)		10	9
Children 0-2 and 6-17 (no 3-5)		14	13
Children 3-17 (no 0-2)		15	14
Not a Parent			-1

ch02	0-1	89-93,99:11-	all
ch05	"	84-93,99:11-	"
ch35	"	89-93,99:11-	"
ch613	"	84-93,99:11-	"
ch1417	"	"	"

Recodes of presence of own child variable. "1" indicates presence of one or more children in that age range, "0" indicates absence of children in the age ranges 0-2, 0-5, 3-5, 6-13, and 14-17 respectively.

a-pfrel prfamrel	pfamrel	0-5	84-	all

Primary family relationship.

	84-93	94-
Not in primary family	0	0
Husband	1	
Wife	2	
Head		1
Spouse		2
Own Child	3	3
Other relative	4	4
Unmarried reference person	5	5

a-vet peafwhen	veteran	1-6	79-	males before 1989 then all

Item 18g. According to BLS documentation, female veterans are not asked this question until 1989. However it appears that at least some females were asked this question prior to 1989 with a reasonable number of females of appropriate age coded with military service in the actual files. When using the military service variable you may want to check to make sure that the ages seem reasonable. A disadvantage of the coding scheme adopted in the CPS is difficulty of adding new wars in consequence, hence no coding for specific post-Vietnam wars.

Vietnam Era (8/64-4/75)	1
Korean War (6/50-1/55)	2
World War II (9/40-7/47)	3
World War I (4/17-11/18)	4
Other Service	5
Non-veteran	6

na gradeat 0-18 79-91 all

Item 18h. Highest grade of school attended. In the original BLS coding for 1979-1988 the value coded for education is one more than the actual grade, so 13 was coded for a person who has at least started the senior year of high school. In 1989-1991 the actual grade is coded, without adding one. So that senior in high school is coded as 12 in the later system. The first edition CD-ROM maintains the BLS coding system, while the second and subsequent editions recode the 1979-1988 values using the later system. **If you are accustomed to the BLS system read this paragraph closely - the old BLS coding is not used here!**

No Schooling	0
1	1
...	
8	8
High School	
9	9
10	10
11	11
12	12
College	
13	13
14	14
15	15
16	16
17	17
18	18

a-hgc gradecp 1-2 79-91 all

Item 18. Was highest graded attended completed?

Yes	1
No	2

a-hga peeduca grade92 31-46 92- all

Item 18h. Highest grade attended. “What is the highest level of school ... has completed or highest degree received?” In 1992 the BLS switched from years of schooling measure to a credential oriented measure. Rumor has it that a labor economist who estimated wage equations for 1991 and 1992 without noticing the difference in the CPS education measure was surprised only by the change in the constant term.

Less than 1st grade	31
1st - 4th grade	32
5th or 6th	33
7th or 8th	34
9th	35
10^{th}	36
11^{th}	37
12^{th} grade NO DIPLOMA	38
High school graduate, diploma or GED	39
Some college but no degree	40
Associate degree -- occupational/vocational	41
Associate degree -- academic program	42
Bachelor's degree (e.g. BA,AB,BS)	43
Master's degree (e.g. MA,MS,MEng,Med,MSW,MBA)	44
Professional school deg. (e.g. MD,DDS,DVM,LLB,JD)	45
Doctorate degree (e.g. PhD, EdD)	46

na relahh 1-6 79-88 all

Item 18b. Relationship to head of household. This is recoded from Relationship to reference person.

Head with other relatives	1
Head with no other relatives	2
Wife of head	3
Other relative of head	4
Non-relative of head with own relatives (includes wife)	5
Non-relative of head with no own relatives	6

a-rrp	relaref	1-10	89-93	all
perrp	relref94	1-12	94-95:2	
	relref95	1-18	95:3-	

Slightly more detail is available for 94 on. In 1995 the partner category is further expanded to distinguish among roommates, partners and boarders. Happily, the additional categories were added at the end, without disturbing existing definitions.

	relaref	relref94	relref95
Head with relative	1	1	1
Head without relative	2	2	2
Husband	3		
Wife	4		
Spouse		3	3
Child	5	4	4
Grandchild		5	5
Parent	6	6	6
Brother/sister	7	7	7
Other relative	8	8	8
Foster child		9	9
Secondary family member:			
Nonrl of hd-own rls in hh	9	10	10
Secondary individual:			
Nonrl of hd-no own rls in hh	10	12	12
Partner/Roommate		11	
Not Used			11
Unmarried partner with relatives			13
Unmarried partner w/o relatives			14
Housemate/rmmate with relatives			15
Housemate/rmmate w/o relatives			16
Roomer/boarder with relatives			17
Roomer/boarder w/o relatives			18

penatvty	penatvty	57-555	94-	all

Country of birth. See codes in labelsYY.do or Appendix G.

pemntvty pemntvty 57-555 94- all

Mother’s country of birth. See codes in labelsYY.do or Appendix G.

pefntvty pefntvty 57-555 94- all

Father’s country of birth. See codes in labelsYY.do or Appendix G.

prcitshp prcitshp 1-5 94- all

Citizenship status.

Native, born in US	1
Native, born in Puerto Rico or US Outlying Area	2
Native, born abroad of American Parent(s)	3
Foreign born, US citizen by naturalization	4
Foreign born, Not a citizen of the US	5

prcitflg prcitflg 0-53 94- all

Citizenship allocation flag. See codes in labelsYY.do.

Unallocated	0
Allocated	1-53

peinusyr	peinusyr	0-13	94-95	prcitshp>1
peinusyr/prinuyer		0-15	96-	prcitshp>1

Immigrant’s year of entry to the United States. “When did ... come to the United States?” Why is this asked of every person every month? Incredibly, BLS has planned for the last few code meanings to change every year! The difference between the first two values is unknown, but may have to do with U.S. possessions. On the CD-ROM NIU is recoded to missing. No “not foreign born” observations were found.

Not in Universe (Born in US)	-1
Not Foreign Born	00
Before 1950	01
1950-1959	02
1960-1964	03
1965-1969	04
1970-1974	05
1975-1979	06
1980-1981	07
1982-1983	08
1984-1985	09
1986-1987	10
1988-1989	11
1990-1991	12
1992-1995	13
Starting January 1996	
1992-1993	13
1994-1997	14
Starting January 1998	
1994-1995	14
1996-1998	15
Starting January 1999	
1996-1999	15
Starting January 2000	
1996-1997	15
1998-2000	16
Starting January 2001	
1998	16
Starting January 2002	
1998-1999	16
2000-2002	17

Wages

Earnings are collected per hour for hourly workers, and per week for other workers. If you want a consistent hourly wage series during entire period, you should use earnwke/uhourse. This gives imputed hourly wage for weekly workers and actual hourly wage for hourly workers. But check earnwke for top-coding. Do not use any wage data that may be present for self-employed workers.

A$hrlywk	paidhr	1-2	79-93	eligible

Unedited Item 25b. “Is ...paid by the hour on this job?” [This job is the current job from uhourse below.]

Yes	1
No	2

a-hrlywk peernhry	paidhre	1-2	79-	eligible

Edited item 25b. “Is ...paid by the hour on this job?” From 1994 on, this question is “HOURLY/NONHOURLY STATUS.”

Yes	1
No	2

a$hrpay	earnhr	0-9999	79-93	paidhr=1

Item 25c. “How much does ...earn per hour?” (in pennies). This is truncated so that when multiplied by usual hours the result is never more than $100,000 per year. Also, in some years a maximum of 9900 is enforced. For 1979 to 1984 earnhr and earnhre are top coded at 99.99. For 1985 on, the top code depends on hours worked and is selected so that earning per hour times usual hours is not more than 1923.07 per week. Examining the data reveals that the top code is not uniformly applied. While there is always a density peak at the top code amount, a similar number of observations are generally present at higher wage rates. Take caution by testing for wages at or above the top code, if appropriate. Tips are not included.

a-herntp	earnhre	0-9999	79-	paidhr=1
prernhly				
pternhly				

Edited Item 25c. “How much does ...earn per hour?” (in pennies) Before 1989 this is always 50 cents or more. Some years this is limited to a range of 50 - 9900. In 1994 a value of 1 cent is converted to missing. The lower bound is 10 cents in 1994 but 20 cents in 1995; 0 cents in 1996+. Top coding is the same as for earnhr.

a$grwek1-4	uearnwk	0-999	79-88	eligible
		0-1999	89-93	

Item 25d. Earnings per week. “How much does...usually earn per week at this job before deductions?” (in dollars) Includes overtime tips and commissions. Use this field (or uearnwke) for hourly workers.

a-brswk	uearnwke	0-1999	79-88	eligible

Edited Item 25d. Earnings per week. How much does...usually earn per week at this job before deductions? Include any overtime pay, commissions, or tips usually received. Dollars. Some with class ‘without pay’ show non-zero earnings. Self-employed should not show earnings, but sometimes do. Source: locations 427-429 on the BLS tape.

a-werntp	earnwke	0-999	79-88	eligible
		0-1923	89-93	
prernwa		0-1923	94-97	
pternwa		0-2884	98-	

Edited or computed earnings per week in this job. Includes overtime tips and commissions. For hourly workers, computed Item 25a times Item 25c appears here. For weekly workers, edited Item 25d appears here. Also for 1989 on, there are no zero values, suggesting an undocumented change in universe. For 1979-1988 this is from locations 417-419.

a%uslhrs pxhrusl1	I25a	0-1,0-8,0-53	79-	eligible
a%hrlywk pxernhry	I25b	"	"	"
a%hrspay prhernal	I25c	0-1,0-8,0-1	79-93,96+	"
a%grswk prwernal	I25d	"	"	paidhr=1

These are allocation flags for the items I25a through I25d. An item may be edited but not allocated, i.e. a correction. In the pre-1989 tapes 'not allocated' is indicated by a missing value indicator. This has been changed to 0 on CD-ROM for consistency with the 1989 on coding. I25a > 0 always means that usual hours are allocated on the CD-ROM in any year. Note that Stata variable names are case sensitive.

For 1979-1988 the coding scheme is:

Not allocated	0
allocated	1

For 1989 to 1993 the coding scheme is:

No change	0
Value to blank	1
Blank to value	2
Value to value	3
Allocated	4
Value to value -- no error	5
Refusal to value, allocated -- no error	6
Blank to NA -- no error	7
Blank to NA -- error	8

I25c never shows a value of 4.

For 1994 and beyond I25a and I25b range from 0 to 53. Values over three signify allocated data. The types of allocations are in labelsYY.do and in an appendix to the CPS documentation. Values between 23 and 33 indicate allocations based on a prior month interview in the same household, other allocations are less reliable.

For 1996 on the coding scheme for I25c and I25d is:

Not allocated	0
allocated	1

Employment

For the employed, current job is the job held in the reference week (the week before the survey). Persons with 2 or more jobs are classified in the job at which they worked the most hours during the reference week. The unemployed are classified according to their latest full time job lasting two weeks or more or by the current job (full or part-time). The industry and occupation questions are also asked of departing rotations (dp) not in the labor force who have worked in the last five years. The universe for I&O is all private workers for pay, as defined by the edited class of worker variable. The universe for class of worker variables is approximately those in the labor force, or who have been in the labor force within the last 5 years (1989-1993). For 1994 onward the universe includes those in the labor force or worked within last year. In some years non-workers may be in the universe only if their past job was full-time.

a$clswkr	class	1-8	79-93
a-clswkr	classer1	1-8	89-93
peio1cow	class94	1-8	94-

Item 23e, class of worker. Class and classer1 have the same coding, a-clswkr is the edited version of a$clswkr. Note that the years of availability are not the same. Class94 has a new coding to distinguish between non-profit and for-profit employment. Other changes are gratuitous. Some ‘without pay’ show earnwke positive.

	class & classer1	class94
Private, for profit	1	4
Private, non-profit	1	5
Federal Government	2	1
State Government	3	2
Local Government	4	3
Self-employed (incorporated)	5	6
Self-employed (not incorporated)	6	7
Without pay	7	8
Never worked or never worked full-time	8	missing

na	classer	1-5	79-88	

Edited and recoded class of worker.

Private	1
Government	2
Self-employed	3
Without pay	4
Never worked or never worked full-time	5

a-row	Classer2	1-7	89-93	all

Edited and recoded a$clswkr. The self employed (incorporated) category seems to have been absorbed into self employed unincorporated. Class94 (above) replaces this variable after 1993 though this variable continues to be available in the source.

Private	1
Federal Government	2
State Government	3
Local Government	4
Self-employed, unincorporated	5
Without pay	6
Never worked	7

na	esr	1-7	79-88	all
a-lfsr	lfsr89	1-7	89-93	
pemlr	lfsr94	1-7	94-	

Employment Status Recode Last week. This is later called the Labor Force Status Recode. A value 0 of undefined meaning occurs in 1989 only. These variables control the universe for many variables in this section. “Without pay” refers to family business or farm.

	esr	lfsr89	lfsr94	
Working	1	1	1	E
With a job, not at work	2	2	2	E
Looking	3	4	4	U
Layoff		3	3	U
Housework	4			NILF
School	5			NILF
Unable to work/Disabled	6		6	NILF
Working without pay		5		NILF
Unavailable for work		6		NILF
Other (Includes Retired)	7	7	5,7	NILF

na	ind70	17-937	79-82	

This is the 3-digit Industry Classification from the 1970 Census. See labelsYY.do or Appendix A for codes. This variable is present on the BLS tape in 1983, but is not to be relied on for that year and is not included in the extracts.

a-ind peio1icd	ind80	10-991	83-	

Item 23b. This is the 3-digit Industry Classification Code from the 1980 Census. See labelsYY.do or Appendix B for codes. The universe is unclear but seems to be all those working or who have worked in the last five years(1983-1988) or last year (1994 onward).

	dind	1-52	79-	

This is an NBER created 2-digit Detailed Industry Classification Code that is consistent over all the years covered. See labelsYY.do or appendix A for codes. The BLS supplied 2-digit industry codes are so inconsistent with 3-digit data that they have been dropped from the CD-ROM extracts.

na	occ70	1-984	79-82	see ind70

This is the 3-digit Occupational Classification from the 1970 Census. ‘What kind of work was ... doing?” This variable is present on the original tape in 1983, but is not to be relied on for that year. See labelsYY.do or Appendix C for codes.

a-occ peio1ocd	occ80	3-905	83-	

This is the 3-digit Occupational Classification from the 1980 Census. ‘What kind of work was ... doing?’ See Appendix D for codes. (/sources/labels/occYY.do is too large for Stata 6.)

na	docc70	0-44	79-82	

This is the 2-digit Detailed Occupation Recode from the 1970 Census. See labelsYY.do or Appendix C for codes. For 1983 the CPS documentation shows a field with this definition, but the contents of the field are inappropriate.

a-dtocc docc80 1-46 83-
prdtocc1

This is the 2-digit Detail Occupation Recode from the 1980 Census. The 1979-1982 3-digit classification would not easily be coded into this form.

a-ernel eligible 1-2 79:5- all
prerelg

Eligibility Flag. This flag marks non-self-employed workers for pay. In the original files "1" always marks a private worker for pay, but the alternative may be "0" or missing, depending on the year. For the CD-ROM these later values are translated to "2" for consistency. Note that this variable starts in mid- 1979.

Earnings eligible	1
other	2

a-majact activlwr 1-8 79-93 all
1-8 89-93

Edited Item 19. “What was...doing most of LAST WEEK?”(Major Activity)

Working	1
With a job	2
Looking for work	3
Keeping house	4
At school	5
Unable to work	6
Retired	7
Other	8

a$majact doinglw 1-8 79-93 all

Unedited and unallocated Item 19. ‘What was...doing most of LAST WEEK?’ Codes are the same as a-majact above.

a-hrs1 hourslwa 0-99 79-93 working

Unedited Item 20a. ‘How many hours did...work last week at all jobs?’

a$uslhrs uhours 0-99 79-93 eligible

Unedited Item 25a. ‘How many hours per week does...USUALLY work at this job?’ (Main job)

a-uslhrs peernhro uhourse 0-99 79- eligible

Edited Item 25a. ‘How many hours per week does...USUALLY work at this job?’ [1989 trough 1993 the range is 1-99.] The allocation flag for this variable is noted with the earnings variables above. For 1994 on the job is the ‘main job’ and the answer ‘hours vary’ is translated to missing in the extracts.

a$uslft uhours35 1-2 79-93 ESR=1&item 20a<35

Unedited and unallocated Item 20c. “Does...USUALLY work 35 hours or more a week at this job?” Part 1.

Blank	missing
Yes	1
No	2

a-hrs1 pehractt hourslw 1-99 79- working

Edited item 20a. “How many hours did...work last week at all jobs?” For 1994 and after this is allowed to go to 198 hours on the original tape. This is truncated on the CD-ROM.

a-ftreas	reasonlw	1-15	79-93	uhourse<34
perhrrsn3	reason94	1-13	94-	hourslw<35

Edited Item 20c. Part 2. 'What is the main reason ... worked less than 35 hours last week?'

	79-93	94-
NIU (89-93)	.	
Slack work	01	1
Material shortage	02	
Plant or machine repair	03	
New job started during week	04	
Job terminated during week	05	
Job started or ended during week		3
Could only find part-time work	06	
Holiday	07	6
Labor dispute	08	9
Bad weather	09	10
Own illness	10	5
On vacation	11	4
Too busy with school, house	12	
Did not want full-time work	13	
Full time work week is <35 hours	14	
Other	15	13
Seasonal work		2
Child care problems		7
Other family / personal obligations		8
School / Training		11
Civic / military duty		12

a-whyabs	absentlw	1-8	79-93	not at work or looking
peabsrsn	absent94	1-14	94-	not at work

Item 21a. "Why was...absent from work last week?" Note that the Universe changes from the labor force to just those with a job for 1994 on.

	absentlw	absent94
Own illness	1	5
On vacation	2	4
Bad weather	3	10
Labor dispute	4	9
New job to begin within 30 days	5	3
Temporary layoff (under 30 days)	6	1
Indefinite layoff (30 days or more)	7	1
Other	8	14
Slack work / business conditions		2
Child care problems		6
Other family / personal		7
Maternity / paternity		8
School / training		11
Civic / military		12
Does not work		13

a$ftreas	why35lw	1-15	79-93	
prptrea	why3594	1-23	94-	uhourse<35

Unedited and unallocated Item 20c. Part 2. 'What is the reason ...USUALLY works less than 35 hours a week?' In 1994 and on full and part-time workers are distinguished. A few persons with hours>34 are included here, reasons unknown.

	why35lw	why3594 FT	why3594 PT
Blank	missing		
Slack work	1	1	14
Material shortage	2		
Plant or machine repair	3		
New job started last week	4	3	
Job terminated during week	5	3	
Could only find part-time work	6		15
Holiday	7	6	
Labor dispute	8	9	
Bad weather	9	10	
Own illness	10	5	19
On vacation	11	4	
Too busy with house, school, etc.	12		
Did not want full-time work	13		
Full-time work week under 35 hours	14		
Other	15	13	23
Seasonal		2	16
Child care problems		7	17
Other family obligations		8	18
School / training		11	20
Civic / military		12	
Social Security limit on earnings			21

na	ftpt79	0-5	79-88	civilians

Full-time or part-time labor force status.

Not in labor force	0
Employed full-time	1
Part-time for economic reasons	2
Unemployed full-time	3
Employed part-time	4
Unemployed part-time	5

a-wkstat	ftpt89	1-7	89-93	all
prwkstat	ftpt94	1-12	94-	94-

It is unknown if these categories fully enumerate the possibilities; ‘economic reasons’ refer to labor demand factors.

	ftpt89	ftpt94
Not in labor force	1	1
Full-time schedule	2	2
PT for economic reasons usually FT	3	3
PT for non-economic reasons usually PT	4	7
PT for economic reasons usually PT	5	6
Unemployed FT	6	11
Unemployed PT	7	12
Not at work usually FT		5
PT for non-economic reasons usually FT		4
FT usually PT for economic reasons		8
FT usually PT for non-economic reasons		9
Not at work usually PT		10

na	ptstat	0,5,6	79-88	all

Part-time status.

All other	0
Voluntary part-time workers	5
Part-time for economic reasons	6

a-ftpt peschft	studftpt	1-2	84-	age 16-24

Item 26b. “Is...enrolled in a school as a full-time or part-time student [this week]?” (There is no documentation for a code for non-students, but they are coded as missing).

Full-time	1
Part-time	2

Union variables

unionmm	unionmm	1-2	84-93	eligible

Item 25E (Unedited). On this job, is... a member of a labor union or an employee association similar to a union? The CPS documentation claims that the universe is all departing rotations, but class<5 (Private or government worker for pay) would seem to be the actual universe.

Yes	1
No	2

a-unmem peernlab	unionmme	1-2	83-	eligible

Item 25E (Edited). On this job, is...a member of a labor union or an employee association similar to a union? The universe is subject to the same comment mentioned under unionmm above.

Yes	1
No	2

a$uncov peerncov	unioncov	1-2	83-	not union members

Item 25F (Edited). On this job, is...covered by a union or employee association contract? (Note universe: What about union members not covered by a contract?)

Yes	1
No	2

Crosswalk CPS Names				
NBER Name		Years	CPS 89-93 Name	CPS 94- Name
Miscellaneous variables				
hhid	Household ID	79-	H-ID	HRHHID
lineno	Personal Line Number	79-	A-LINENO	PULINENO
minsamp	Month in sample	79-	H-MIS	HRMIS
hrlonglk	Longitudinal Link Indicator	94-		HRLONGLK
year		79-	H-YEAR	H-YEAR
intmonth	Interview month	79-	H-MONTH	HRMONTH
hhnum	Household number	79-	H-HHNUM	HUHHNUM
weight	Final Weight x 100	79-	A-FNLWGT	PWSSWGT
earnwt	Earnings weight for all races	79-	A-ERNLWT	PWORWGT
Geography				
state	State	79-	HG-ST6O	GESTCEN
smsastat	SMSA status code	79-	H-METSTA	GEMETSTA
centcity	Central city status	79-95:5, 95:9-	HG-MSAS	G(E/T)MSAST
smsa70	SMSA size	75-85:6		
smsa80	"	85:10-94:3	HG-MSSZ	GEMSASZ
smsa93	"	95:9-		GECMSASZ
smsarank	SMSA ranking	79-85:6		
pmsarank	PMSA ranking	86-94:3	HG-PMSA	GEPMSRK
cmsarank	CMSA/MSA ranking	86-94:3	HG-MSAR	GEMSARK
icntcity	Individual central city code	86-	H-INDVCC	GEINDVCC
msafips	MSA/PMSA FIPS code	89-	HG-MSAC	G(E/T)MSA
cmsacode	CMSA code	89-94:3, 95:9-	HG-CMSA	G(E/T)CMSA
Demography				
sex	Sex	79-	A-SEX	PESEX
race	Race	79-	A-RACE	PERACE
ethnic	Ethnicity	79-	A-REORGN	PRORIGIN
age	Age	79-	A-AGE	PEAGE
marital	Marital Status	79-	A-MARITL	PRMARSTA
ownchild	Number of own children in pri. fam.	84-93,99:11-	A-PFNOCD	PRNMCHLD
chldpres	Presence of own children < 18 in pf	84-93,99:11-	A-PFPRCD	PRCHLD
ch02	Presence of own children 0-2 in p. f.	89-93,99:11-	--	--
ch05	Presence of own children 0-5 in p. f.	84-93,99:11-	--	--
ch35	Presence of own children 3-5 in p. f.	84-93,99:11-	--	--
ch613	Presence of own children 6-13 in p.f.	84-93,99:11-	--	--
ch1417	Presence of own children 14-17 in pf	84-93,99:11-	--	--
pfamrel	Primary family relationship	84-	A-PFREL	PRFAMREL
veteran	Veteran	79-	A-VET	PEAFWHEN
gradeat	Highest grade attended	79-91	A-HGA	
gradecp	Whether completed highest grade	79-91	A-HGC	

grade92	Highest grade attended	92-	A-HGA	PEEDUCA
relahh	Relationship to household head	79-88		
relaref	Relationships to reference person	89-93	A-RRP	
relref94	“	94-95:2		PERRP
relref95	“	95:3-		“
penatvty	Country of birth	94-		PENATVTY
pemntvty	Mother s country of birth	94-		PEMNTVTY
pefntvty	Father s country of birth	94-		PEFNTVTY
prcitshp	Citizenship status	94-		PRCITSHP
prcitflg	Citizen allocation flag	94-		PRCITFLG
peinusyr	Immigrant s year of entry	94-		PEINUSYR/PRINUYER
Wages				
paidhr	Paid by the hour	79-93	A$HRLYWK	
paidhre	“	79-	A-HRLYWK	PEERNHRY
earnhr	Earnings per hour	79-93	A$HRPAY	
earnhre	“	79-	A-HERNTP	P(R/T)ERNHLY
uearnwk	Earnings per week	79-93	A$GRWEK1	
uearnwke	“	79-88	A-BRSWK	
earnwke	“	79-	A-WERNTP	P(R/T)ERNWA
I25a	“ : paid by hour	79-	A%USLHRS	PXHRUSL1
I25b	“ : paid by hour	79-	A%HRLYWK	PXERNHRY
I25c	“ : earnings per hour	79-93, 96+	A%HRSPAY	PRHERNAL
I25d	“ : usual earnings per hour	79-93, 96+	A%GRSWK	PRWERNAL
Employment				
class	Class of worker	79-93	A$CLSWKR	
classer1	“	89-93	A-CLSWKR	
class94	“	94-		PEIO1COW
classer	“	79-88		
classer2	“	89-93	A-RCOW	PRCOW1
esr	Employment status recode	79-88		
lfsr89	“	89-93	A-LFSR	
lfsr94	“	94-		PEMLR
ind70	3-digit industry code	79-82		
ind80	“	83-	A-IND	PEIO1ICD
dind	“	79-		
occ70	3-digit occupation code	79-82		
occ80	“	83-	A-OCC	PEIO1OCD
docc70	2-digit occupation code	79-82		
docc80	“	83-	A-DTOCC	PRDTOCC1
eligible	Eligibility flag	79:5-	A-ERNEL	PRERELG
activlwr	Major activity last week	79-93	A-MAJACT	
doinglw	What was doing most last week	79-93	A$MAJACT	
hourslwa	Usual hours	79-93	A-HRS1	
uhours	“	79-93	A$USLHRS	
uhourse	“	79-	A-USLHRS	PEHRUSL1

uhours35	Usually works >=35 hrs at this job	79-93	A$USLFT	
hourslw	How many hrs last week all jobs	79-	A-HRS1	PEHRACTT
reasonlw	Reason <=35 hours last week	79-93	A-FTREAS	
reason94	“	94-		PEHRRSN3
absentlw	Why absent from work last week?	79-93	A-WHYABS	
absent94	“	94-		PEABSRSN
why35lw	Why not at least 35 hrs last week	79-93	A$FTREAS	
why3594	“	94-		PRPTREA
ftpt79	Full-time or part-time status	79-88		
ftpt89	“	79-93	A-WKSTAT	
ftpt94	“	94-		PRWKSTAT
ptstat	Part-time status	79-88		
studftpt	Enrolled as a student full/part time	84-	A$FTPT	PESCHFT
Union Status				
unionmm	Union member	83-93	A$UNMEM	
unionmme	“	83-	A-UNMEM	PEERNLAB
unioncov	Covered by a union contract	83-	A$UNCOV	PEERNCOV