

Monthly Current Population Survey Public Use Microdata Files

Public use microdata files from the Current Population Survey (CPS) are freely available from the Census Bureau, enabling researchers to conduct their own analyses with the data. This paper describes the files and how to obtain them, as well as discussing the data dictionaries and presenting some example calculations.

Confidentiality of monthly CPS public use microdata files

The monthly CPS public use files contain information collected in the CPS about individual respondents and members of their households. The files have one record for each household member, and each file contains responses from a particular month. These files do not contain any personally identifiable information; all such information has been removed to protect respondent confidentiality. Because of this, specific variables or particular response options for some variables may not appear on the files. For example, detailed geographic information does not appear for all households. Additionally, some variables are topcoded to protect confidentiality—that is, there is an upper bound for certain variables. For example, the age variable (PRTAGE) for a 101-year-old would be topcoded at 85 for the public use file.

Beginning in January 2011, the Census Bureau incorporated additional safeguards in the monthly CPS public use files to ensure that respondent-identifying information is not disclosed. In general, respondents' ages were altered, or "perturbed", in the public use microdata files to further protect the confidentiality of survey respondents and the data they supply. One result of the measures taken to enhance data confidentiality is that labor force and other estimates from the monthly public use microdata files may no longer exactly match estimates published by the Bureau of Labor Statistics (BLS); the BLS estimates are based on internal files that have not been perturbed. Although certain topside labor force estimates will continue to match published data—such as the overall levels of employed, unemployed, and not in the labor force—estimates below the topside level (for example, employment status by age, sex, race, and ethnicity) all have the chance of differing slightly from the published data.

In addition, estimates calculated using characteristics such as industry, occupation, hours worked, and duration of unemployment, along with all other characteristics not expressly listed above, are subject to such differences. All such differences should fall well within the sampling variability associated with CPS estimates.

Obtaining CPS public use microdata files

Monthly CPS public use microdata files from January 1994 through the present can be downloaded free of charge from the Census Bureau's DataWeb FTP page, at http://thedataweb.rm.census.gov/ftp/cps_ftp.html. Also available on this website are periodic extract files with additional information, such as files containing the 2015 and 2016 variables on certifications and licenses. Data dictionaries for all files can also be downloaded.

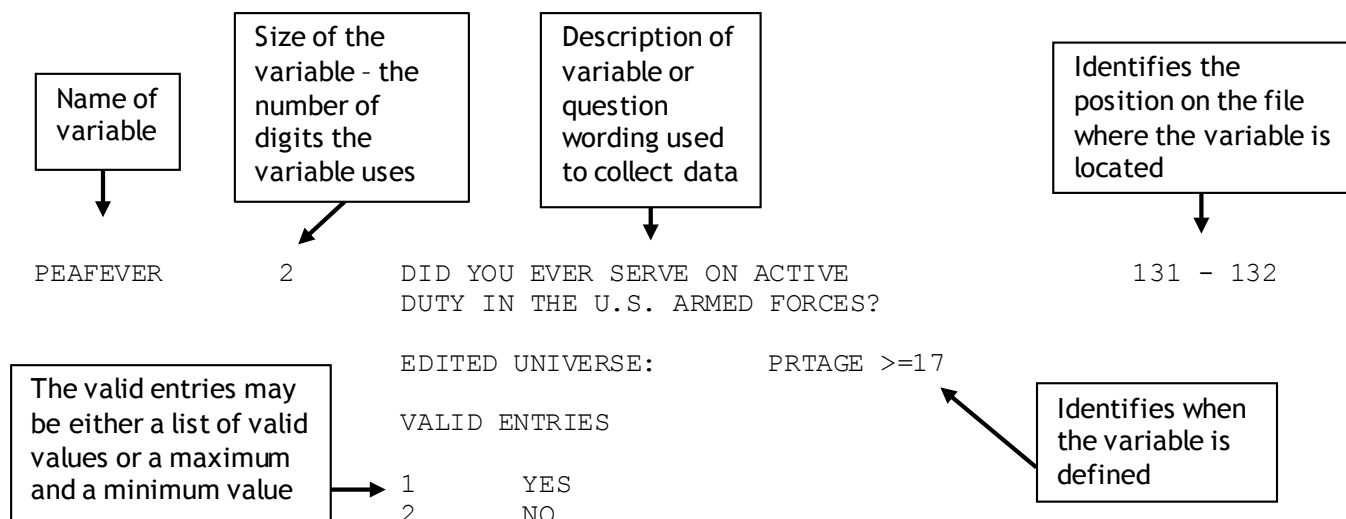
The data files are available in compressed format, as either .gz or .zip files. Some are also available as compressed UNIX files. Once uncompressed, the data are in a fixed-width .dat file that a statistical software program can read. The data dictionary provides the number of characters and the location for each variable.

There are other organizations that maintain copies of the CPS public use microdata files, some of which have more historical files available. Please note that variable names may differ on files obtained from a source other than the Census Bureau.

Data dictionaries for monthly CPS public use files

Along with the monthly CPS public use files, data dictionaries can be downloaded from the Census FTP DataWeb page (http://thedataweb.rm.census.gov/ftp/cps_ftp.html). These will provide extensive information about each variable included in the data files.

Below is a sample entry from the monthly CPS public use data dictionary:



Naming conventions and definitions for monthly CPS public use files

Variables on the monthly CPS public use files are named according to specific rules (the only two exceptions to these rules are the variables `OCCURNUM` and `QSTNUM`):

- The first character of the variable name identifies the level of the variable (whether it refers to a person, a household, or a household location).
- The second character of the variable name identifies the variable type (whether it is unedited, edited, a recode, an allocation flag, topcoded, or a statistical weight).
- The remaining characters consist of a descriptive name.

The following table gives more information about the first character of the variable name.

First Character	Level of Variable	Definition
G	Geography variable	Geography variables are collected in CPS and identify a household's location. For example, <code>GESTFIPS</code> gives the FIPS state code for the location of the household. Geography variables have separate values for each household in the monthly CPS public use files. Thus, they are the same for each member of a single household.
H	Household variable	Household variables are collected in CPS and describe a household characteristic. For example, <code>HRNUMHOU</code> gives the total number of people living in the household. Household variables have separate values for each household in the monthly CPS public use files. Thus, they are the same for each member of a single household.
P	Person variable	Person variables collected in CPS describe a characteristic of a person. For example, <code>PRTAGE</code> gives the age of an individual. Person variables have separate values for each person in the monthly CPS public use files.

The following table gives more information about the second character of the variable name.

Second character	Variable Type	Definition
U	Unedited Variable	<p>An unedited variable generally is produced by the Computer Assisted Telephone Interview (CATI) or Computer Assisted Personal Interview (CAPI) instrument and is either collected or assigned during the interview.</p> <p>“U” variables on monthly CPS public use files are either person (“P”) or household (“H”) variables.</p>
E	Edited Variable	<p>An edited variable has gone through an editing process (a process checking for consistency and allocating missing values). Values of edited variables are almost always equal to values of the corresponding unedited variables. Data differ when a value is allocated or imputed by the processing system based on CPS allocation rules. Allocations are typically performed when the unedited variable contains a missing value or a value of "don't know" or "refused," though they may also be done when the respondent provides contradictory information.</p> <p>An edited version of a variable exists only if that variable goes through an editing process. If there are no edits for a variable, then only an unedited version of the variable exists.</p> <p>“E” variables on the monthly CPS public use files can be either person (“P”), household (“H”), or geography (“G”) variables.</p>
R	Recode	<p>A recode is a variable calculated by the processing system from a combination of other items on the file. For example, PRMJOC1 is the major occupation code for a household member’s main job; this is not a response to a question but rather a variable that summarizes (or “groups”) the more finely detailed occupation variable PEIO1OCD.</p> <p>Recodes on the CPS file are either person (“P”) or household (“H”) variables.</p>
T	Topcode Indicator	<p>These variables indicate topcoding, or the assigning of maximum values. Four topcode variables on the CPS file are indicator flags that relate to earnings or age, and the others are variables that have been topcoded. There are a few variables with topcoded data that do not contain “T” in the second character (such as PRERNWA and PRERNHLY).</p> <p>Topcode Indicator variables on the CPS file can be either person (“P”) variables or geography variables (“G”).</p>
W	Statistical weight	<p>These variables are pre-calculated statistical weights for use in producing accurate estimates. See the section below on “CPS Statistical Weights”.</p> <p>“W” variables on the CPS files are either person (“P”) or household (“H”).</p>
X	Allocation Flag	<p>Each edited person or household variable has a corresponding allocation flag indicating the nature of the allocation. For example, if PUSEX is blank, PESEX would be allocated and have the corresponding allocation flag of PXSEX = 41 (blank to allocated value). See the section on allocation flags for the complete list of values. There are three “X” variables that do not have corresponding “E” variables but rather correspond to “R” or “T” variables (PXAGE, PXRACE1, and PXINUSYR).</p> <p>“X” variables on the monthly CPS public use files can be either person (“P”) or household (“H”) variables.</p>

Using these rules, variables can be more readily understood based on their names. For example, the variable PESEX can be broken down as follows:

- The first character “P” indicates that this is a person-level variable that was collected or created through the CPS interviews.
- The second character “E” indicates that this variable went through an editing process; it also means that a corresponding allocation flag, PXSEX, will indicate the nature of the allocation.
- The final part of the variable name, “SEX,” is descriptive.

Some questions asked in the CPS interviews allow for more than one response. For such multiple-entry questions, there is a separate variable for each possible response. Each variable has the same descriptive name but a different (sequential) number. For example, respondents can provide up to six answers to the question “You said you have been trying to find work—how did you go about looking?” The variable names are PULKDK1, PULKDK2, PULKDK3, etc.

Not all CPS variables are included in the public use files. When there is an edited variable, the corresponding unedited variable is usually omitted from the files. This is typically done to protect the confidentiality of CPS respondents as required by law. If an unedited variable is included on the files, it generally means that an edited version does not exist. As with all other variables included on the public use files, publicly available unedited variables cannot be used to identify individual respondents.

Valid entries

Almost all variables on the monthly CPS public use files have a number of valid entries or a range of valid values. For example, the variable PESEX has two valid entries: 1 for male and 2 for female. The variable PRTAGE, on the other hand, has a range of valid values—any entry between 0 and 85 (except 81 through 84) is considered valid. Individual valid values or a range of valid values are listed under each variable in the data dictionary. A few variables have so many valid values that they are not included in the data dictionary; instead, they are provided in an appendix or a separate document. (References to these are included as a “Note” under the relevant variables in the data dictionary.) One example of such a variable is PEIO1ICD, which identifies the industry code of the respondent’s main job.

Many variables have the following possible valid values:

Value	Description
-1	Blank
-2	Don’t know
-3	Refused

Since so many variables have these possible values, they are not shown as valid entries for each item.

A few variables on the monthly CPS public use files do not list either valid values or a range of valid values. These are typically identifying variables, such as the primary identifiers for CPS records (HRHHID and HRHHID2).

Allocation flags

Item nonresponse refers to a missing variable in an otherwise completed questionnaire. Generally, this occurs when respondents either don’t know the answer to a question or they refuse to answer, but it can occur for other reasons as well. For example, a variable may not be recorded due to an interviewer or computer error. Item nonresponse should not be ignored because it is unlikely to occur at random. Ignoring missing data and restricting analysis to records with reported values relies on the implicit (and possibly inaccurate) assumption that all respondents are equally likely or unlikely to respond to the item and that estimates are approximately unbiased.

For virtually every person-level or household-level edited item (“PE” and “HE” variables), there is a corresponding allocation flag whose second character is “X.” All remaining characters of the two variables’ names are the same. For example, PXSEX is the allocation flag for PESEX.

Generally, allocation flags have the following list of possible values:

- 0 Value - no change
- 1 Blank - no change
- 2 Don’t know - no change
- 3 Refused - no change
- 10 Value to value
- 11 Blank to value
- 12 Don’t know to value
- 13 Refused to value
- 20 Value to longitudinal value
- 21 Blank to longitudinal value
- 22 Don’t know to longitudinal value
- 23 Refused to longitudinal value
- 30 Value to allocated longitudinal value (currently not used)
- 31 Blank to allocated longitudinal value (currently not used)
- 32 Don’t know to allocated longitudinal value (currently not used)
- 33 Refused to allocated longitudinal value (currently not used)
- 40 Value to allocated value
- 41 Blank to allocated value
- 42 Don’t know to allocated value
- 43 Refused to allocated value
- 50 Value to blank
- 52 Don’t know to blank
- 53 Refused to blank

Each digit of these valid values identifies how and why edited variables were allocated. The first digit indicates how the allocation was made to the “E” (or edited) variable.

First Digit	Meaning
0 or Blank	No change between “U” variable and “E” variable
1	“E” variable changed to a value
2	“E” variable changed to a longitudinal value (the corresponding value from a prior CPS interview)
3	“E” variable changed to an allocated longitudinal value (the corresponding allocated value from a prior CPS interview)—currently not used
4	“E” variable changed to allocated value
5	“E” variable changed to a blank

The second digit indicates why the “U” variable was allocated—that is, whether the value was changed, missing, don’t know, or refused.

Second Digit	Meaning
0	“U” variable was equal to some value
1	“U” variable was blank (or -1)
2	“U” variable was don’t know (or -2)
3	“U” variable was refused (or -3)

All “X” variables follow this standard format. However, there are a few variables that are not “X” variables but do indicate whether a variable has been allocated, and these flags typically have

different valid entries. Examples include PRCITFLG (allocation flag for PRCITSHP), PRERNWAL (allocation flag for PRERNWA), and PRHERNAL (allocation flag for PRERNHLY).

All “PE” and “HE” variables have a corresponding “X” variable. However, there are occasional “X” variables that correspond to an “R” or “T” variable rather than an “E” variable. These include PXRACE1 (allocation flag for PTDRACE), PXAGE (allocation flag for PRTAGE), and PXINUSYR (allocation flag for PRINUYER).

Edited universe

Edited variables and recodes are defined for certain universes, and these are listed in the data dictionary. For example, PEMJNUM (number of jobs) is only defined for those who have indicated that they have more than one job. Therefore, the universe for PEMJNUM is PEMJOT = 1 (PEMJOT is the multiple job status of the individual; a value of 1 indicates that the individual has more than one job).

Certain variables might initially appear to be the same because their descriptions are very similar. These variables are different in that they were asked of different groups of survey respondents. For example, the variables PEERNH1O and PEERNH2 both have the same question text of “Excluding overtime pay, tips, and commissions, what is your hourly rate of pay on your main job?” The difference between these two variables has to do with which group of respondents was asked each question, and this can be determined by looking at the edited universe. PEERNH1O was asked of respondents with PEERNPER = 1, or those who said it was easiest to report their earnings hourly. PEERNH2, on the other hand, was asked of respondents who said it was easiest to report their earnings some other way than hourly even though they were paid hourly.

CPS statistical weights

The Census Bureau weights each response in order to estimate aggregate population totals. The public use data files therefore have several variables that are used as weights in different calculations. All of these variables have an implied four decimals, though the decimal itself is not included. Hence, each weight must be divided by 10,000 when used.

BLS published estimates almost all are calculated using the composited final weight. This weight is controlled to population counts of the civilian noninstitutional population age 16 and over, as well as to labor force status for selected demographic groups (age, sex, race, and ethnicity) and geography (state). However, some estimates are calculated using the outgoing rotation weight or the veterans weight. For more information about the weighting process, see Chapter 10 of *Design and Methodology: Current Population Survey, Technical Paper 66* (<https://www.census.gov/prod/2006pubs/tp-66.pdf>).

Weight Name	Weight Definition	Weight Description
PWFMWGT	Family Weight	Used for estimates of families.
PWLGWGT	Longitudinal Weight	Used for adult records matched from month-to-month. (Note that this is not the weight used to produce BLS estimates of labor force flows; those estimates use a weight calculated by BLS that is not publicly available.)
PWORWGT	Outgoing Rotation Weight	Used for computing estimates of information collected only in the outgoing rotation groups (that is, people who have been in the sample for either their 4 th or their 8 th month). Examples of such information include weekly earnings and union membership status.
PWSSWGT	Second Stage Weight	The final step before creating the composited final weight for major labor force statistics. Also used in creating other weights, including the veterans weight and the household weight. It is the most demographically correct weight.

Weight Name	Weight Definition	Weight Description
PWVETWGT	Veterans Weight	Used when computing estimates of veterans and nonveterans. Controlled to population estimates maintained by the Department of Veterans Affairs.
PWCMPWGT	Composited Final Weight	Used to create BLS published labor force statistics. Only created for the civilian noninstitutional population 16 and over.
HWHHWGT	Household Weight	Used when estimating household characteristics.

Linking the CPS public use files

Each CPS public use file covers one month. While most researchers use the files as a point-in-time snapshot, either using one month or combining multiple months, some researchers wish to leverage the longitudinal aspect of the CPS to investigate how data for the same individuals change over time. In order to link the same individuals across months, three variables must be used: HRHHID, HRHHID2, and PULINENO.

Frequently-used variables

There are a number of variables that are commonly used by researchers. These define critical demographic concepts, such as race, sex, and age, as well as important details of a person's labor force status. Researchers also often rely on geographic variables that identify household location. The table below includes some of the more commonly used variables.

Characteristic	Variable Name	Notes
<i>Demographic Variables</i>		
Type of person	PRPERTYP	Identifies whether the person is a child, adult, or serving in the U.S. Armed Forces
Age	PRTAGE	Called PEAGE before January 2015
Sex	PESEX	
Race	PTDTRACE	
Hispanic ethnicity	PEHSPNON	
Detailed Hispanic ethnicity	PRDTHSP	
Educational attainment	PEEDUCA	
Certification and licensing status	PECERT1	
Disability status	PRDISFLG	
Veteran status	PEAFEVER	
Marital status	PEMARITL	
Foreign-born status	PENATVTY	
<i>Labor Market Variables</i>		
Labor force status	PEMLR	Defines employed, unemployed, and not in the labor force
Full- or part-time work status	PRWKSTAT	Identifies full- and part-time status as defined by actual hours of work. Also identifies usual and actual full- and part-time status by economic and noneconomic reasons. (Can also be used to identify employed, unemployed, and not in the labor force.)
Weekly earnings	PRERNWA	Calculated for all workers. Collected for one-quarter of the sample.
Hourly earnings	PRERNHLY	Only calculated for hourly-paid workers. Collected for one-quarter of the sample.

Characteristic	Variable Name	Notes
Multiple job status	PRSJMJ	
Usual weekly hours worked	PEHRUSLT PEHRUSL1 PEHRUSL2	PEHRUSLT describes all jobs combined, PEHRUSL1 describes the main job, and PEHRUSL2 describes the other job(s).
Actual hours worked during the survey reference week	PEHRACTT PEHRACT1 PEHRACT2	PEHRACTT describes all jobs combined, PEHRACT1 describes the main job, and PEHRACT2 describes the other job(s). The survey reference week is generally the week containing the 12 th of the month.
Occupation (main job)	PEIO1OCD PRDTOCC1 PRMJOC1 PRMJOCGR	These variables have varying levels of detail. PEIO1OCD has the most detailed occupation breakout, PRDTOCC1 has 22 categories, PRMJOC1 has 10 categories, and PRMJOCGR has 6 categories. Occupation of second job is available in PEIO2OCD.
Class of worker (main job)	PEIO1COW PRCOW1 PRDTCOW1	Class of worker for second job is available in PEIO2COW, PRCOW2, and PRDTCOW2.
Industry (main job)	PEIO1ICD PRDTIND1 PRIMIND1 PRMJIND1	These variables have varying levels of detail. PEIO1ICD has the most detailed industry breakout, PRDTIND1 has 51 categories, PRDTIND1 has 21 categories, PRIMIND1 has 21 categories, and PRMJIND1 has 13 categories. Industry of second job is available in PEIO2ICD.
Geographic Variables		
State	GESTFIPS	
Metropolitan Statistical Area	GTCBSA	
County	GTCO	

Selected BLS definitions

BLS defines a number of labor force concepts using the variables mentioned above. Some of the most widely-used concepts are listed in the following table.

Concept	Variable definition
Civilian Noninstitutional Population	PRPERTYP = 2 and PRTAGE ≥ 16
Unemployed	3 ≤ PEMLR ≤ 4
Employed	1 ≤ PEMLR ≤ 2
Labor force	1 ≤ PEMLR ≤ 4
At work part time for economic reasons	PRWKSTAT = 3 or 6

Generating estimates using the monthly CPS public use files

When generating estimates from the CPS data files, follow these steps:

- **Identify the file(s) needed:** Identify the time period of interest and the file(s) that should be used.
- **Identify the variables needed:** There are several key pieces of documentation to use when working with the data files that can help you determine which variables are best for your purpose. The data dictionaries and technical documentation available on the Census DataWeb FTP site include variable definitions.

- **Identify the appropriate weight:** As discussed above, BLS generally uses the composited final weight for its labor force calculations.
- **Perform the calculation:** Make sure the formula you are using is correct.
- **Verify your results (if possible):** It is always helpful to verify at least one estimate with an official CPS estimate. Note that estimates generated through the public use files—other than such high-level estimates as total employment, unemployment, and not in the labor force—may not exactly match official estimates due to age perturbation done to protect respondent confidentiality. CPS labor force estimates are published monthly in the Employment Situation news release, available from the BLS website (<https://www.bls.gov/bls/news-release/empst.htm>). Many other monthly, annual, and quarterly tables are available at <https://www.bls.gov/cps/>, and many estimates are also available through the BLS online LABSTAT database (<https://www.bls.gov/cps/data.htm>).

Reliability of estimates

The public use data allow researchers to create estimates for any number of subgroups of the population. However, it is important to keep in mind that the more stratified the subgroups are, the smaller the sample size becomes. Estimates based on very small samples are considerably less likely to be reliable.

Examples

Example 1: Calculate the number of unemployed people in January 2015.

- Identify the file(s) needed: This calculation uses the January 2015 public use file.
- Identify the variables needed:
 - PEMLR will define the unemployed
 - Unemployed: $3 \leq \text{PEMLR} \leq 4$
- Identify the statistical weight needed: Like virtually all BLS labor force estimates, this uses the composited final weight, PWCMPWGT. Statistical weights on the public use files have four implied decimal places, so PWCMPWGT should be divided by 10,000 when used. (Note that the use of the composited final weight automatically restricts the calculation to those age 16 and over who are not members of the U.S. Armed Forces.)
- Perform the calculation:

The number of unemployed people is given by

$$\text{Unemployment} = \sum_i \text{wgt}_i * U_i$$

Here U_i is an indicator variable that equals 1 if person i is unemployed; otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

- Verify the estimate: The number should match the overall *not seasonally adjusted* unemployment estimate published by BLS for January 2015 in the Employment Situation news release (https://www.bls.gov/news.release/archives/empst_02062015.pdf).

Example 2: Calculate the number of people in the labor force in January 2015.

People in the labor force include all people who are either employed or unemployed.

- Identify the file(s) needed: This calculation uses the January 2015 public use file.

- Identify the variables needed:
 - PEMLR will define the labor force (employed and unemployed)
 - Labor force: $1 \leq \text{PEMLR} \leq 4$
- Identify the statistical weight needed: Like virtually all BLS labor force estimates, this uses the composited final weight, PWCMPWGT. Statistical weights on the file have four implied decimal places, so PWCMPWGT should be divided by 10,000 when used. (Note that the use of the composited final weight automatically restricts the calculation to those age 16 and over who are not members of the U.S. Armed Forces.)
- Perform the calculation:

The number of people in the labor force is given by

$$\text{Labor force} = \sum_i wgt_i * L_i$$

Here L_i is an indicator variable that equals 1 if person i is in the labor force (either employed or unemployed); otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

- Verify the estimate: The number should match the overall *not seasonally adjusted* labor force estimate published by BLS for January 2015 in the Employment Situation news release (https://www.bls.gov/news.release/archives/empst_02062015.pdf).

Example 3: Calculate the national unemployment rate for January 2015.

The unemployment rate is the unemployed as a percentage of the labor force.

- Identify the file(s) needed: This calculation uses the January 2015 public use file.
- Identify the variables needed:
 - PEMLR will define both the unemployed and the labor force
 - Unemployed: $3 \leq \text{PEMLR} \leq 4$
 - Labor force: $1 \leq \text{PEMLR} \leq 4$
- Identify the statistical weight needed: Like virtually all BLS labor force estimates, this uses the composited final weight, PWCMPWGT. Statistical weights on the file have four implied decimal places, so PWCMPWGT should be divided by 10,000 when used. (Note that the use of the composited final weight automatically restricts the calculation to those age 16 and over who are not members of the U.S. Armed Forces.)
- Perform the calculation:

The unemployment rate is given by

$$\text{Unemployment Rate} = \frac{\sum_i wgt_i * U_i}{\sum_i wgt_i * L_i} * 100$$

Here U_i is an indicator variable that equals 1 if person i is unemployed; otherwise, it is 0. Similarly, L_i is an indicator variable that equals 1 if person i is in the labor force; otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

- Verify the estimate: The number should match the overall *not seasonally adjusted* unemployment rate published by BLS for January 2015 in the Employment Situation news release (https://www.bls.gov/news.release/archives/empst_02062015.pdf).

Example 4: Calculate the unemployment rate for Hispanics age 16 to 19 in January 2015.

- Identify the file(s) needed: This calculation uses the January 2015 public use file.
- Identify the variables needed:
 - PEMLR will define both the unemployed and the labor force
 - Unemployed: $3 \leq \text{PEMLR} \leq 4$
 - Labor force: $1 \leq \text{PEMLR} \leq 4$
 - PEHSPNPN will define whether a person is Hispanic
 - Hispanic: $\text{PERHSPNON} = 1$
 - PRTAGE will define whether a person is aged 16 to 19
 - Age 16-19: $16 \leq \text{PRTAGE} \leq 19$
- Identify the statistical weight needed: Like virtually all BLS labor force estimates, this uses the composited final weight, PWCMPWGT. Statistical weights on the file have four implied decimal places, so PWCMPWGT should be divided by 10,000 when used. (Note that the use of the composited final weight automatically restricts the calculation to those age 16 and over who are not members of the U.S. Armed Forces.)
- Perform the calculation:

The unemployment rate is given by

$$\text{Unemployment Rate} = \frac{\sum_i \text{wgt}_i * U_i}{\sum_i \text{wgt}_i * L_i} * 100$$

Here U_i is an indicator variable that equals 1 if person i is unemployed *and* is 16 to 19 *and* is Hispanic; otherwise, it is 0. Similarly, L_i is an indicator variable that equals 1 if person i is in the labor force *and* is 16 to 19 *and* is Hispanic; otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

- Verify the estimate: As discussed earlier, statistics calculated using the public use files may not exactly match estimates published by BLS. This is due to the age perturbation that is done for the public use files to protect respondent confidentiality. The number should be similar to the overall *not seasonally adjusted* unemployment rate published by BLS for January 2015 in the Employment Situation news release (https://www.bls.gov/news.release/archives/empst_02062015.pdf).

Example 5: Calculate the unemployment level and the unemployment rate for Hispanics age 16 to 19 in 2015.

- Identify the file(s) needed: This calculation uses 12 public use files, January 2015 through December 2015.

- Identify the variables needed:
 - PEMLR will define both the unemployed and the labor force
 - Unemployed: $3 \leq \text{PEMLR} \leq 4$
 - Labor force: $1 \leq \text{PEMLR} \leq 4$
 - PEHSPNPN will define whether a person is Hispanic
 - Hispanic: $\text{PERHSPNON} = 1$
 - PRTAGE will define whether a person is age 16 to 19
 - Age 16 to 19: $16 \leq \text{PRTAGE} \leq 19$
- Identify the statistical weight needed: Like virtually all BLS labor force estimates, this uses the composited final weight, PWCMPWGT. Statistical weights on the file have four implied decimal places, so PWCMPWGT should be divided by 10,000 when used. (Note that the use of the composited final weight automatically restricts the calculation to those age 16 and over who are not members of the U.S. Armed Forces.)
- Perform the calculation:
To get the average monthly number of unemployed 16- to 19-year-old Hispanics in 2015, we need to aggregate all individual records for the year and divide by 12. The easiest way to do to this is to create one file that contains all 12 months of data. (An annual average estimate is *not* calculated by computing the levels for each month individually and then averaging them.) The unemployment estimate is calculated using the following formula:

$$\text{Unemployment} = \frac{1}{12} * \sum_i \text{wgt}_i * U_i$$

Here U_i is an indicator variable that equals 1 if person i is unemployed *and* is Hispanic *and* is aged 16 to 19 (regardless of which month the observation occurred in); otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

In order to calculate the unemployment rate, it is necessary to also calculate the number of 16- to 19-year-old Hispanics in the labor force. The labor force estimate is calculated using the same method as that used for the unemployment estimate. Hence:

$$\text{Labor force} = \frac{1}{12} * \sum_i \text{wgt}_i * L_i$$

Here L_i is an indicator variable that equals 1 if person i is in the labor force *and* is Hispanic *and* is aged 16 to 19; otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the composited final weight associated with person i .

Finally, the unemployment rate for the year is the number of unemployed 16- to 19-year-old Hispanics in 2015 divided by the number of 16- to 19-year-old Hispanics in the labor force in 2015 and then multiplied by 100. Using the unemployment and labor force levels calculated above,

$$\text{Unemployment Rate} = \frac{\text{Unemployment}}{\text{Labor Force}} * 100$$

Verify the estimate: As discussed earlier, statistics calculated using the public use files may not exactly match estimates published by BLS. This is due to the age perturbation that is done for the public use files to protect respondent confidentiality. The number should be similar to the estimate for Hispanics age 16 to 19 in table 3 of the BLS report *Labor Force Characteristics by Race and Ethnicity, 2015* (<https://www.bls.gov/opub/reports/race-and-ethnicity/2015/pdf/home.pdf>).

Example 6: Calculate the employment-population ratio of veterans for 2015.

The employment-population ratio is defined as the number of employed as a percentage of the civilian noninstitutional population. The civilian noninstitutional population includes all people age 16 and over who are not in the Armed Forces or in institutions such as correctional facilities, nursing homes, or mental health care facilities. Veterans are those who had served on active duty in the Armed Forces but were not on active duty at the time of the survey. BLS typically limits estimates for veterans to those who are age 18 and over. Thus, the 2015 employment-population ratio for veterans would be the number of employed veterans age 18 and over in 2015 as a percentage of the total number of veterans age 18 and over in the civilian noninstitutional population in 2015.

- Identify the file(s) needed: This calculation uses 12 public use files, January 2015 through December 2015.
- Identify the needed variables:
 - PEMLR will define the employed
 - Employed: $1 \leq \text{PEMLR} \leq 2$
 - PRPERTYP will define the civilian noninstitutional population
 - Civilian noninstitutional population: $\text{PRPERTYP} = 2$
 - PEAFEVER will define each person's veteran status
 - Veteran: $\text{PEAFEVER} = 1$
 - PRTAGE will identify whether a person is age 18 or over
 - Age 18 and over: $18 \leq \text{PRTAGE}$
- Identify the needed statistical weight: Here the weight will be the veterans weight, PWVETWGT. Statistical weights on the file have four implied decimal places, so PWVETWGT should be divided by 10,000 when used.
- Perform the Calculation:

To get the average monthly number of employed veterans age 18 and over in 2015, we need to aggregate all individual records for the year and divide by 12. The easiest way to do this is to create one file that contains all 12 months of data. (An annual average estimate is *not* calculated by computing the levels for each month individually and then averaging them.) The employment level is calculated using the following formula:

$$\text{Employment} = \frac{1}{12} * \sum_i \text{wgt}_i * E_i$$

Here E_i is an indicator variable that equals 1 if person i is employed *and* is a veteran *and* is age 18 or over (regardless of which month the observation occurred in); otherwise, it is 0. The wgt_i variable is the statistical weight given to person i , which, in this example, should be the veterans weight associated with person i .

In order to calculate the employment-population ratio, it is necessary to also calculate the average monthly number of veterans age 18 or over in the civilian noninstitutional population. The civilian noninstitutional population estimate is calculated using the same method as that used for the employment estimate. Hence:

$$\text{Civilian noninstitutional population} = \frac{1}{12} * \sum_i \text{wgt}_i * P_i$$

Here P_i is an indicator variable that equals 1 if person i is in the civilian noninstitutional population *and* is a veteran *and* is age 18 or over; otherwise, it is 0. The wgt_i variable is the

statistical weight given to person i , which, in this example, should be the veterans weight associated with person i .

The employment-population ratio is equal to the number of employed veterans age 18 and over divided by the total number of veterans age 18 and over in the civilian noninstitutional population and then multiplied by 100. Using the employment and civilian noninstitutional population estimates calculated above:

$$\text{Employment} - \text{population ratio} = \frac{\text{Employment}}{\text{Civilian noninstitutional population}} * 100$$

- Verify the estimate: As discussed earlier, statistics calculated using the public use file may not exactly match estimates published by BLS. This is due to the age perturbation that is done for the public use files to protect respondent confidentiality. The number should be similar to the overall employment-population ratio for veterans that is published by BLS in the Employment Situation of Veterans 2015 news release (https://www.bls.gov/news.release/archives/vet_03222016.pdf).