

# Calculating Approximate Standard Errors and Confidence Intervals for Current Population Survey Estimates

This document provides information about calculating approximate standard errors for estimates from the Current Population Survey (CPS). It also includes examples of how confidence intervals for estimates can be calculated. A November 2018 update of this document introduces  $\alpha$  and  $\beta$  parameters and a slightly different methodology for calculating standard errors.

The CPS sample size is designed to meet a specified reliability criteria for unemployment estimates at the national and state level. The requirement is that, assuming a national unemployment rate of 6.0 percent, an over-the month change of roughly 0.2 percentage points be statistically significant at the 90-percent level of confidence.

A table showing changes in selected labor force indicators with statistical significance tests is online at <https://www.bls.gov/cps/documentation.htm#reliability>. The table is updated each month with the publication of the *Employment Situation* news release.

Approximate standard errors and confidence intervals for CPS estimates can be calculated using instructions in this document and the parameter and factor tables (PF-1 through PF-16) available online at <https://www.bls.gov/cps/parameters-and-factors-for-calculating-standard-errors.xlsx>. (These tables mirror tables A-1 through A-16 of the monthly *Employment Situation* news release.) The parameter and factor tables allow users to calculate approximate standard errors for a wide range of estimated levels, rates, and percentages, and also changes over time. The parameters and factors are used in formulas that are commonly called generalized variance functions.

## **Reliability of Current Population Survey Estimates**

An estimate based on a sample survey like the CPS has two types of error—*sampling error* and *nonsampling error*. The estimated standard errors provided in this publication are approximations of the true sampling errors. They incorporate the effect of some nonsampling errors in response and enumeration, but do not account for any systematic biases in the data.

## **Nonsampling error**

Nonsampling error refers to errors due to factors that are not related to sample selection. These errors can be attributed to many sources, and many would affect results of a census as well as a survey.

These sources include:

- The sampling frame may have imperfections. For example, the frame may exclude some housing units. Other housing units may be included in error or appear more than once. This can cause undercoverage or overcoverage.
- Household rosters may inadvertently exclude persons or include persons. Also, based on scope definitions by age or other criteria, errors may cause a person to be inadvertently excluded or included. This may cause undercoverage or overcoverage that can vary, particularly by demographic characteristics.
- Questions may not be designed properly, though extensive testing makes this a minor issue for CPS.
- Responses to some items may be incorrect or may be missing entirely. For example, questions may be misinterpreted, or a person giving information about another may not have sufficient knowledge to provide correct information.
- Housing unit nonresponse may introduce bias. No response is obtained from about 10% of the housing units determined to be eligible (in scope) for a number of reasons. For example, some housing units refuse to cooperate, and no successful contact is made with others. If responding households have different characteristics than nonresponding households, the estimates may be biased.
- Statistical methods to fill in for missing data or to correct for housing unit nonresponse are imperfect.
- Ratio estimation (used in second-stage weighting and composite weighting) is known to introduce small biases.
- CPS composite estimation is affected by month-in-sample (MIS) bias. One trait of MIS bias is that the estimated unemployment rate for a panel (or rotation group) is highest when it is first included in the CPS.
- There can also be errors made by interviewers, even falsification. These errors are minimized by thorough training, use of laptops for administering interviews, routine quality control checks, and a reinterview program.
- Data transmission and processing errors are possible, but should be eliminated by CPS quality control procedures.

The full extent of nonsampling error in the CPS is unknown. The effect is small on estimates of relative change, such as month-to-month change. The effect is small on means and rates, particularly unemployment rates. Estimates of monthly levels tend to be affected to a greater degree. More information can be found in Current Population Survey Design and Methodology (Technical paper 66, October 2006) online at <https://www.census.gov/prod/2006pubs/tp-66.pdf>.

## Sampling error and confidence intervals

When a sample, rather than the entire population, is surveyed, estimates differ from the true population values that they represent. The component of this difference that occurs because samples differ by chance is known as sampling error, and its variability is measured by the standard error of the estimate. A sample estimate and its estimated standard error can be used to construct confidence intervals; when these estimates are unbiased, the statistical properties of confidence interval “coverage” are known.

The following are examples of confidence intervals:

- A 90% confidence interval is the range from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate. The true population value is unknown, but there is an approximate 90% probability that the interval includes or “covers” the true population value.
- A 95% confidence interval is the range from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate. The true population value is unknown, but there is an approximate 95% probability that the interval includes or “covers” the true population value.

Confidence interval statements like these are approximately true for the CPS, and this is especially so for rates and changes over time. For a more complete explanation of confidence interval coverage refer to a standard survey methodology text, such as chapter 1.7 in the 3<sup>rd</sup> edition of *Sampling Techniques* by William G Cochran (Wiley, 1977).

When examining differences between estimates, the question often asked is: are the two estimates significantly different? The difference may be for the same characteristic at two different time periods: is the change significantly different? The difference may be between different subpopulations in the same time period, such as the unemployment rate for men versus the unemployment rate for women. Given an unbiased estimate of difference/change and an unbiased estimate of standard error for that difference/change, confidence intervals can be constructed and statements such as the following can be made:

- If zero lies *outside of* the 90% confidence interval from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate, the difference/change is said to be significantly different at the 90% level of confidence. If zero lies *in* the interval, the difference/change is said to be *not* significantly different at the 90% level of confidence.
- If zero lies *outside of* the 95% confidence interval from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate, the difference/change is said to be significantly different at the 95% level of confidence. If zero lies *in* the interval, the difference/change is said to be *not* significantly different at the 95% level of confidence.

Approximate standard errors and confidence intervals for CPS estimates can be calculated using the instructions below and the parameter and factor tables (PF-1 through PF-16) available online at <https://www.bls.gov/cps/parameters-and-factors-for-calculating-standard-errors.xlsx>. (These tables mirror tables A-1 through A-16 of the monthly *Employment Situation* news release.) The parameter and factor tables allow users to calculate approximate standard errors for a wide range

of estimated levels, rates, and percentages, and also changes over time. The parameters and factors are used in formulas that are commonly called generalized variance functions.

The approximate standard errors calculated using the parameter and factor tables (PF-1 through PF-16) are based on the sample design and estimation procedures as of 2015, and reflect the population levels and sample size as of that year. Guidance for calculating standard error estimates for historical CPS data may be found in the “Reliability of the estimates” section of the Household Data Technical Notes (*Employment and Earnings*, February 2006, pages 12-19 or printed pages 193-200) online at [https://www.bls.gov/cps/eetech\\_methods.pdf](https://www.bls.gov/cps/eetech_methods.pdf).

Information presented here may be of use to researchers working with official BLS estimates as well as those computing estimates using the public use CPS microdata files, available from the Census Bureau's FTP site ([https://thedataweb.rm.census.gov/ftp/cps\\_ftp.html](https://thedataweb.rm.census.gov/ftp/cps_ftp.html)) or from their DataFerrett tool (<https://dataferrett.census.gov/>). Note that estimates generated using the public use microdata files, except for a few topside estimates, will not match official BLS estimates.<sup>1</sup>

### **Calculating approximate standard errors and confidence intervals**

A table showing changes in selected labor force indicators with statistical significance tests is online at <https://www.bls.gov/cps/documentation.htm#reliability>.

### **Using parameter and factor tables PF-1 through PF-16**

These tables give  $\alpha$  and  $\beta$  parameters that can be used with formulas to calculate approximate monthly standard errors for a wide range of estimated levels, proportions, and rates. Factors are provided to convert monthly measures into approximate standard errors of estimates for other periods (quarterly and yearly averages) and approximate standard errors for changes over time (consecutive monthly changes, changes in consecutive quarterly and yearly averages, and changes in monthly estimates 1 year apart).

The standard errors for estimated changes in level estimates from one time period to the next (for example, one month to the next or one year to the next) depend more on the monthly levels than on the size of the changes. Likewise, the standard errors for changes in rates (or percentages)

---

<sup>1</sup> Beginning with data for January 2011, the Census Bureau incorporated additional safeguards in the CPS public use microdata 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 public use microdata files will no longer exactly match most estimates published by BLS, which are based on internal, nonpublic-use files. Although certain topside labor force estimates for all persons 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 (such as employment status by age, sex, race, and Hispanic or Latino ethnicity) all have the chance of differing slightly from the published data. In addition, estimates calculated using characteristics such as industry, occupation, hours worked, 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.

depend more on the monthly rates (or percentages) than on the size of the changes. Accordingly, the factors presented in tables PF-1 through PF-16 are applied to the monthly standard error approximations for levels, percentages, or rates; the magnitudes of the changes do not come into play. Factors are not given for estimated changes between nonconsecutive months (except for changes of monthly estimates 1 year apart); however, the standard errors may be assumed to be higher than the standard errors for consecutive monthly changes.

**IMPORTANT:** CPS data are published as levels in thousands, so 4 million unemployed people would be displayed as 4,000 in a table. Similarly, an entry of 2,460 in a CPS table represents 2,460,000 people. When calculating standard errors and confidence intervals, you must use the actual level at full precision, not the data as published in thousands.

**Standard errors of estimated levels**

The approximate standard error  $se(x)$  of  $x$ , an estimated monthly level, can be obtained using the formula below, where  $\alpha$  and  $\beta$  are the parameters from parameter and factor tables PF-1 through PF-16 associated with a particular characteristic, and  $N$  is the total civilian noninstitutional population 16 years and over.

$$se(x; N) = \sqrt{(\alpha + \beta N) \left( x - \frac{x^2}{N} \right)}$$

Note that  $x$  is the monthly level (not in thousands).

*Illustration – single month level*

Assume that, in a given a month, there are an estimated 4 million unemployed men, and the civilian noninstitutional population is 250 million. Obtain the appropriate  $\alpha$  and  $\beta$  parameters from table PF-1 (Men, 16 years and over; Unemployed). Use the formula for  $se(x; N)$  to compute an approximate standard error on the estimate of  $x = 4,000,000$  where  $N = 250,000,000$ .

$$\begin{aligned} \alpha &= 1050.17 \\ \beta &= 0.00000883 \end{aligned}$$

$$\begin{aligned} &se(4,000,000; 250,000,000) \\ &= \sqrt{(1050.17 + (0.00000883 * 250,000,000)) * \left( 4,000,000 - \left( \frac{4,000,000^2}{250,000,000} \right) \right)} = 113,235 \end{aligned}$$

$$se(4,000,000; 250,000,000) \approx 113,000$$

**Standard errors of estimated levels for quarterly or annual averages or changes over time**

Tables PF-1 through PF-16 provide factors that can be used to compute approximate standard errors of levels for other time periods or for changes over time. For each characteristic, factors (*f*) are given for:

- Consecutive month-to-month changes
- Changes in monthly estimates 1 year apart
- Quarterly averages
- Changes in consecutive quarterly averages
- Yearly averages
- Changes in consecutive yearly averages

For a given characteristic, the correct factor from tables PF-1 through PF-16 is used in the following formula, which also uses the  $\alpha$  and  $\beta$  parameters from the same line of the table. A three-step procedure for using the formula is given. The *f* in the formula is frequently called an *adjustment factor*, because it appears to adjust a monthly standard error  $se(x)$ . However, the  $x$  and  $N$  in the formula are not monthly levels, but averages of monthly levels (see examples listed below).

$$se(x; N; f) = f * se(x; N) = f * \sqrt{(\alpha + \beta N)(x - \frac{x^2}{N})}$$

Note that  $x$  and  $N$  are averages of monthly levels over the designated period.

Step 1. Average monthly levels appropriately in order to obtain  $x$ . Levels for 3 months are averaged for quarterly averages, and those for 12 months are averaged for yearly averages. For changes in consecutive levels, average over the 2 months, 2 quarters, or 2 years involved. For changes in monthly estimates 1 year apart, average the 2 months involved.

Step 2. Calculate an approximate standard error  $se(x)$ , treating the average  $x$  from step 1 as if it were an estimate of level for a single month. For a given characteristic, obtain parameters  $\alpha$  and  $\beta$  from the applicable PF table.

Step 3. Determine the standard error  $se(x; f)$  on the average level or on the change in level. Multiply the result from step 2 by the appropriate factor  $f$ . The  $\alpha$  and  $\beta$  parameters used in step 2 and the factor  $f$  used in this step come from the same line in the corresponding PF table.

*Illustration – consecutive month change in level*

Continuing the previous example, suppose that in the next month the estimated number of unemployed men increases by 150,000, from 4,000,000 to 4,150,000. Also suppose the civilian noninstitutional population 16 years and over increases by 200,000, from 250,000,000 to 250,200,000.

Step 1. The averages of the two monthly levels for the estimate and the total population are  $x = 4,075,000$  and  $N = 250,100,000$ .

Step 2. Apply the  $\alpha$  and  $\beta$  parameters from table PF-1 (Men, age 16 years and over; Unemployed) to the average  $x$  and  $N$ , treating them like an estimate and a population value for a single month.

$$\begin{aligned}\alpha &= 1050.17 \\ \beta &= 0.00000883\end{aligned}$$

$$\begin{aligned}se(4,075,000; 250,100,000) \\ &= \sqrt{(1050.17 + (0.00000883 * 250,100,000)) * (4,075,000 - \frac{4,075,000^2}{250,100,000})} \\ &= 114,290\end{aligned}$$

$$se(4,075,000; 250,100,000) \approx 114,000$$

Step 3. Obtain  $f = 1.10$  from the same row of table PF-1 in the column “Consecutive month-to-month change,” and multiply the factor by the result from step 2 to calculate the standard error for the change between the 2 months.

$$se(150,000) = f * se(4,075,000; 250,100,000) = 1.10 * 114,290 = 125,719$$

$$se(150,000) \approx 126,000$$

For an approximate 90-percent confidence interval, compute  $1.645 * 125,719 = 206,808 \approx 207,000$ . Next, subtract the number from and add the number to 150,000 to obtain an interval of -57,000 to 357,000. This is an approximate 90-percent confidence interval for the true change, and since this interval includes zero, one cannot assert at this level of confidence that any real change has occurred in the unemployment level. The result also can be expressed by saying that the apparent change of 150,000 is not statistically significant at a 90-percent confidence level.

*Illustration – quarterly average level*

Suppose that an approximate standard error is desired for a quarterly average of the Black or African American employment level. Suppose that the estimated employment levels for the 3 months making up the quarter are 14,900,000; 15,000,000; and 15,100,000, and the corresponding levels of the civilian noninstitutional population 16 years and over are 249,800,000; 250,000,000; and 250,200,000.

Step 1. The averages of the three monthly levels are  $x = 15,000,000$  and  $N = 250,000,000$ .

Step 2. Apply the  $\alpha$  and  $\beta$  parameters from table PF-2 (Black or African American; Total; Employed) to the average  $x$  and  $N$ , treating them like an estimate for a single month.

$$\alpha = -592.49$$

$$\beta = 0.00000816$$

$$se(15,000,000; 250,000,000)$$

$$= \sqrt{(-592.49 + (0.00000816 * 250,000,000)) * (15,000,000 - \frac{15,000,000^2}{250,000,000})} = 142,863$$

$$se(15,000,000; 250,000,000) \approx 143,000$$

Step 3. Obtain  $f = 0.86$  from the same row of table PF-2 in the column “Quarterly averages,” and multiply the factor by the result from step 2 to calculate the standard error of the quarterly average.

$$se(15,000,000) = 0.86 * 142,863 = 122,862$$

$$se(15,000,000) \approx 123,000$$

*Illustration – consecutive quarter change in level*

Continuing the example, suppose that, in the next quarter, the estimated average employment level for Blacks is 15,400,000, based on monthly levels of 15,300,000; 15,400,000; and 15,500,000. This is an estimated increase of 400,000 over the previous quarter. Also suppose the average population level in the next quarter is 250,600,000, based on monthly levels of 250,400,000; 250,600,000; and 250,800,000.

Step 1. The average of the two quarterly levels (15,000,000 and 15,400,000) is  $x = 15,200,000$ . The average of the two quarterly population levels (250,000,000 and 250,600,000) is  $N = 250,300,000$ .

Step 2. Apply the  $\alpha$  and  $\beta$  parameters from table PF-2 (Black or African American; Total; Employed) to the average  $x$  and  $N$ , treating them like an estimate for a single month.

$$a = -592.49$$

$$b = 0.00000816$$

$$se(15,200,000; 250,300,000)$$

$$= \sqrt{(-592.49 + (0.00000816 * 250,300,000)) * (15,200,000 - \frac{15,200,000^2}{250,300,000})} = 143,878$$

$$se(15,200,000; 250,300,000) \approx 144,000$$

Step 3. Obtain  $f = 0.79$  from the same row of table PF-2 in the column “Change in consecutive quarterly averages,” and multiply the factor by the result from step 2 to calculate the standard error of the change in quarterly averages.

$$se(400,000) = 0.79 * se(15,200,000) = 0.79 * 143,878 = 113,664$$

$$se(400,000) \approx 114,000$$

For an approximate 95-percent confidence interval, compute  $1.96 * 113,664 = 222,781 \approx 223,000$ . Subtract the number from and add the number to 400,000 to obtain an interval of 177,000 to 623,000. The interval excludes zero. Another way of stating this is to observe that the estimated change of 400,000 clearly exceeds 1.96 standard errors, or 223,000. One can conclude from these data that the change in quarterly averages is statistically significant at a 95-percent confidence level.

### Standard errors of estimated rates, ratios, and percentages

As shown in the formula below, the approximate standard error  $se(p;y)$  of an estimated rate or percentage  $p$  depends, in part, upon the number of persons  $y$  in its base or denominator. Generally, rates and percentages are not published unless the monthly base is greater than 75,000 persons, the quarterly average base is greater than 60,000 persons, or the yearly average base is greater than 35,000 persons. The  $\alpha$  and  $\beta$  parameters are obtained from the correct PF table. When the base  $y$  and the numerator of  $p$  are from different categories within the table, use the  $\alpha$  and  $\beta$  parameters from the PF table relevant to the numerator of the rate or percentage.

$$se(p; y) = \sqrt{\frac{(\alpha + \beta y)}{y} p(100 - p)}$$

Note that  $y$  (not in thousands) is the base of percent  $p$ , and  $se(p; y)$  is in percent.

#### *Illustration – single month percent*

For a given month, suppose  $y = 156,000,000$  employed persons. Of this total, 27,000,000, or  $p = 17.3$  percent, are classified as part-time workers. Obtain the  $\alpha$  and  $\beta$  parameters from table PF-9 (Full- or part-time status; Part-time workers) that are relevant to the numerator of the percentage. Apply the formula to obtain:

$$\alpha = -1636.59$$

$$\beta = 0.00002042$$

$$se(17.3; 156,000,000) = \sqrt{\frac{(-1636.59 + (0.00002042 * 156,000,000))}{156,000,000} * 17.3 * (100 - 17.3)}$$

$$= 0.119 \approx 0.1 \text{ percent}$$

For an approximate 95-percent confidence interval, compute  $1.96 * 0.119 = 0.233$  percent  $\approx 0.2$  percent. Subtract this from and add this to the estimate of  $p = 17.3$  percent to obtain an approximate confidence interval of 17.1 percent to 17.5 percent.

### **Standard errors of percentages for quarterly or annual averages or changes over time**

Factors from tables PF-1 through PF-16 can be used to compute approximate standard errors on rates, ratios, and percentages for other periods or for changes over time. As with levels, there are three steps in the procedure for using the formula.

$$se(p; y; f) = f * se(p; y) = f * \sqrt{\frac{(\alpha + \beta y)}{y} p(100 - p)}$$

Note that  $p$  and  $y$  are averages of monthly estimates over the designated period, and  $se(p; y; f)$  is in percent.

Step 1. Appropriately average estimates of monthly rates or percentages to obtain  $p$ , and also average estimates of monthly levels to obtain  $y$ . For changes in consecutive averages, average over the 2 months, 2 quarters, or 2 years involved. For changes in monthly estimates 1 year apart, average the 2 months involved.

Step 2. Calculate an approximate standard error  $se(p; y)$ , treating the averages  $p$  and  $y$  from step 1 as if they were estimates for a single month. Obtain the  $\alpha$  and  $\beta$  parameters from the PF table that apply to the numerator of the rate or percentage.

Step 3. Determine the standard error  $se(p; y; f)$  on the average percentage or on the change in percentage. Multiply the result from step 2 by the appropriate factor  $f$ . The  $\alpha$  and  $\beta$  parameters used in step 2 and the factor  $f$  used in this step come from the same line in the appropriate PF table.

#### *Illustration – consecutive month change in percentage*

Continuing the previous example, suppose that, in the next month, there are 156,600,000 employed persons, and that 28,000,000, or 17.9 percent, are part-time workers.

Step 1. The month-to-month change is 0.6 percentage point—that is, the share of the employed who worked part time changed from 17.3 percent to 17.9 percent over the month. The average of the two monthly percentages of 17.3 percent and 17.9 percent is needed ( $p = 17.6$  percent), as is the average of the two bases of 156,000,000 and 156,600,000 ( $y = 156,300,000$ ).

Step 2. Apply the  $\alpha$  and  $\beta$  parameters from table PF-9 (Full- or part-time status; Part-time workers) to the averaged  $p$  and  $y$ , treating the averages like estimates for a single month.

$$\alpha = -1636.59$$

$$\beta = 0.00002042$$

$$\begin{aligned}
 se(17.6; 156,300,000) &= \sqrt{\frac{(-1636.59 + (0.00002042 * 156,300,000))}{156,300,000} * (17.6) * (100 - 17.6)} \\
 &= 0.120 \approx 0.1 \text{ percent}
 \end{aligned}$$

Step 3. Obtain  $f = 0.99$  from the same row of table PF-9 in the column “Consecutive month-to-month change,” and multiply the factor by the result from step 2.

$$se(0.6 \text{ percent}) = 0.99 * 0.120 = 0.119 \text{ percent} \approx 0.1 \text{ percent}$$

For an approximate 95-percent confidence interval, compute  $1.96 * 0.119 = 0.233 \approx 0.2$  percent. Subtract this from and add this to the 0.6-percentage point estimate of change to obtain an interval of 0.4 percent to 0.8 percent. Because this interval does not include zero, it can be concluded the change is statistically significant at a 95-percent confidence level.

### **Parameter and factor tables PF-1 through PF-16**

An Excel file with the parameter and factor tables is available online at <https://www.bls.gov/cps/parameters-and-factors-for-calculating-standard-errors.xlsx>.

These PF tables mirror tables A-1 through A-16 of the monthly *Employment Situation* news release.

A table showing changes in selected labor force indicators with statistical significance tests is updated monthly at <https://www.bls.gov/cps/documentation.htm#reliability>.