Change Estimates

This document provides data users with a basic understanding of the estimation methodology and the accuracy of the data for differences (changes) across years. It is also meant to give the user an understanding of how the estimates of difference were calculated. For more information on the sampling, estimation, variance estimation, and confidentiality of the single year estimates, please see the appropriate "Accuracy of the Data" document.

Differences of Estimates Between Two Years

We used two methodologies to compute the difference between two years, based on the type of estimate we are comparing. For most counts, such as of persons, housing units, and households, we identified the "universe" that the estimate belongs to. For example, the number of persons who are Asian alone was considered to be a subset of the total population. We computed the percentage of the universe that the group "Asian alone" represents for each year, calculated the percentage difference between the two years, and computed the margin of error of that difference.

For many of universe lines, as well as for estimates which are not counts (such as medians, aggregates, ratios, and estimates that were already percentages), the difference between the two estimates and the margin of error of the difference were computed.

For this documentation, initial year refers to the first year and final year refers to the last year (or current year).

Difference in Percent Distributions

Let $\hat{X}$ be the estimate of interest, and $\hat{Y}$ be the estimate of the size of the universe. Then,

$$\hat{P} = \frac{\hat{X}}{\hat{Y}} \quad \text{and} \quad SE(\hat{P}) = \frac{1}{\hat{Y}}\sqrt{[SE(\hat{X})]^2 - \hat{P}^2 \times [SE(\hat{Y})]^2}$$

There is a chance that the value under the square root sign in the formula above could be negative. If so, then we set the standard error of the proportion to 0.001.

We compute the proportion and its standard error for both the initial and final years. The difference and its standard error are:

$$DIFF = 100\% \times \left(\hat{P}_{\text{final year}} - \hat{P}_{\text{initial year}}\right)$$

$$SE(DIFF) = 100\% \times \sqrt{[SE(\hat{P}_{\text{final year}})]^2 + [SE(\hat{P}_{\text{inital year}})]^2}$$

Finally, we calculate the 90 percent margin of error of the difference (ME(DIFF)), and round it to the nearest 0.1%.

$$ME(DIFF) = 1.65 \times SE(DIFF)$$

Special Cases:

- If the DIFF or the SE(DIFF) is not zero but is small enough that it would round to zero, we set it to 0.1%.
- If the universe estimate, $\hat{Y}$, is zero for either the initial or final year, then we set DIFF = '-' and ME(DIFF) = '**'.
- If both estimates and both their universes are controlled, then we compute DIFF as above, but we set ME(DIFF) = '*****'.

Difference for Other Estimates

For all other estimates - these can be identified by a "100%" or an "(N/A)" in the percent distribution columns of the data product - the computation is similar to the procedure above:

$$DIFF = \hat{X}_{\text{final year}} - \hat{X}_{\text{initial year}}$$

$$SE(DIFF) = \sqrt{[SE(\hat{X}_{\text{final year}})]^2 + [SE(\hat{X}_{\text{initial year}})]^2}$$

$$ME(DIFF) = 1.65 \times SE(DIFF)$$

Special Cases:

- If the DIFF or the SE(DIFF) is not zero but is small enough that it would round to zero, we set it to the smallest allowable value greater than zero.
- If the standard error for either the initial year or final year is equal to '*' then we set ME(DIFF) = '*'.
- If the standard error for either the initial year or final year is equal to '**' then we set DIFF = '-' and ME(DIFF) = '**'.
- If the standard error for either the initial year or final year is equal to '***' then we set DIFF = '- - -' and ME(DIFF) = '***'.

- If the standard error for both the initial year or final year is equal to '*****' then we set ME(DIFF) = '*****'.

Statistical Significance

The final column in the data product states whether the difference is statistically significant at the 90% confidence level. To determine this, we let

lower bound = DIFF - ME(DIFF)

upper bound = DIFF + ME(DIFF)

Now:

- If either the lower bound or upper bound is equal to zero, then the difference *is not* statistically significant.
- If the lower bound is negative and the upper bound is positive, then the difference *is not* statistically significant.
- If both the lower and upper bounds have the same sign (that is both are positive or both are negative), then the difference *is* statistically significant.
- If ME(DIFF) has been set to '*', '**', '***', or '*****', then the significance cannot be computed, and the significance test result is set to '- -'.

Examples - Calculating Differences and Standard Errors of the Difference

We will present some examples based on the real data from 2000 and 2001 to demonstrate the use of the formulas.

Example 1: Difference in Percent Distributions

Let's consider the estimated number of males, 15 years and over, never married in Bronx County, New York, which is in the second table of the profiles. Its universe is the estimated number of males, 15 years and over.

For the initial year, 2000, the estimated number of males, 15 years and over, never married in Bronx County, New York is 184,498, with a lower bound of 180,739 and an upper bound of 188,257. The corresponding estimate of the universe is 422,636, with a lower bound of 422,154 and an upper bound of 423,118.

$\hat{X} = 184{,}498 \qquad \hat{Y} = 422{,}636 \qquad \hat{P}_{\text{initial year}} = 0.437$

Standard Error = (upper bound - estimate) / 1.65

Calculating the standard errors using the upper bounds we have:

SE(184,498) = (188,257 - 184,498) / 1.65 = 2,278

SE(422,636) = (423,118 - 422,636) / 1.65 = 292

Using the formula above for the standard error of the proportion,

$$\text{SE}(0.437) = \frac{1}{422{,}636}\sqrt{2{,}278^2 - 0.437^2 \times 292^2} = 0.005$$

Using the same methods with the final year, 2001, estimates, we obtain

$\hat{P}_{\text{final year}} = 0.442$ and SE(0.442) = 0.007

Now we have the information to calculate the difference and its standard error.

DIFF = 100% × (0.442 - 0.437) = 0.5%

$$\text{SE(DIFF)} = 100\% \times \sqrt{0.005^2 + 0.007^2} = 0.9\%$$

The margin of error and the lower and upper confidence bounds are

ME(DIFF) = 1.65 × 0.9% = 1.5%

lower bound = 0.5% - 1.5% = -1.0%

upper bound = 0.5% + 1.5% = 2.0%

Since the lower bound and upper bound have different signs, the difference is not siginificant at the 90 percent confidence level.

Example 2: Difference for Other Estimates

Let’s consider the estimate of the mean travel time to work in Bronx County, New York, which is in the third table of the profiles. The estimate for the initial year is 40.0, with an upper bound of 40.7.

Standard Error = (upper bound - estimate) / 1.65

SE(40.0) = (40.7 - 40.0) / 1.65 = 0.4

The final year estimate is 41.0 with an upper bound of 41.7, and we can similarly derive its standard error to be 0.4.

The difference and its standard error are:

DIFF = 41.0 - 40.0 = 1.0

SE(DIFF) = $\sqrt{0.4^2 + 0.4^2}$ = 0.6

The margin of error and the lower and upper confidence bounds are

ME(DIFF) = 1.65 × 0.6 = 1.0

lower bound = 1.0 - 1.0 = 0.0

upper bound = 1.0 + 1.0 = 2.0

Since the lower bound is equal to zero, the difference is not siginificant at the 90 percent confidence level.