How to approximate the Standard Error of an estimate over a collection of tracts

This document provides a detailed and fairly complex set of instructions for how to compute the standard error for an estimate made by summing over the published estimates for multiple geographic areas (for example, tracts). Words in all capital letters represent variables on the ACS-Census 2000 downloadable datasets.

Notation

Let P_A and P_C be the ACS and Census proportions respectively for the combined grouping of tracts. Let X(i) denote the value of X for tract i.

$Num_A(i)$ = AVGEST
$Den_A(i)$ = ACS_UNI
$Num_C(i)$ = EST_C2K
$Den_C(i)$ = C2K_UNI

$$P_X = \frac{Num_X(1) + \ldots + Num_X(n)}{Den_X(1) + \ldots + Den_X(n)} = \frac{Num_X}{Den_X} \text{ (where X is either A or C)}$$

Goal: We want to find the standard error of the change in the aggregated grouping of tracts. There will be separate instructions for lines that are initially counts, totals, and ratios.

Approximating the Standard Error - Counts

Most of the 364 lines in the profiles are counts which are converted to proportions before estimating the change. [These are the lines with the variable METHOD equal to ‘P’.]

Obtaining $SE(P_A - P_C)$ is essentially a four step process. Separate methods, with their own special cases, are used to get $SE(P_A)$ and $SE(P_C)$. Then, their values are checked against a maximum value, and finally $SE(P_A - P_C)$ is computed.

Note: ACS_UNI and ACS_UNI_SE are not correctly rounded on your files. These should be rounded to the nearest unit before proceeding.

1. First, to calculate $SE(P_C)$,

$$SE(P_C) = DEFF00 \times \sqrt{\frac{5}{Den_C} \times P_C \times (1 - P_C)} \qquad (1)$$

   If $P_C < 0.02$ then use $P_C = 0.02$, and if $P_C > 0.98$ then use $P_C = 0.98$. Use these capped values only for $SE(P_C)$, not when calculating $P_A - P_C$.

DEFF00 is the Census 2000 long form design factor for the appropriate characteristic. If the tracts you are combining have different values of DEFF00, then use the largest value of DEFF00 in your calculations.

2. Second, to calculate SE(P_A), first calculate

$$SE(Num_A) = \sqrt{\sum_{i=1}^{n} \left(SE\left(Num_A(i)\right)\right)^2} = \sqrt{\sum_{i=1}^{n} \left(AVGSE(i)\right)^2} \qquad (2)$$

$$SE(Den_A) = \sqrt{\sum_{i=1}^{n} \left(SE\left(Den_A(i)\right)\right)^2} = \sqrt{\sum_{i=1}^{n} \left(ACS\_UNI\_SE(i)\right)^2} \qquad (3)$$

Now,

$$SE(P_A) = \frac{1}{Den_A} \sqrt{\left(SE(Num_A)\right)^2 - \left(P_A\right)^2 \times \left(SE(Den_A)\right)^2} \qquad (4)$$

There are two special cases. If $P_A = 1$, then use $P_A = 0$ instead. If the value under the square root sign is negative, then use

$$SE(P_A) = DEFF00 \times \sqrt{\frac{ACSAVWGT - 1}{Den_A} \times P_A \times (1 - P_A)} \qquad (5)$$

In this special case *only*, if $P_A < 0.02$ then use $P_A = 0.02$, and if $P_A > 0.98$ then use $P_A = 0.98$. Use these capped values only for SE(P_A), not when calculating P_A - P_C.

ACSAVWGT is the average of the three single-year county-level average weights.

3. Third, before we calculate the standard error of the change, if either SE(P_A) or SE(P_C) is greater than 0.7, then set that SE to 0.7.

4. And finally, fourth,

$$SE(P_A - P_C) \doteq \sqrt{\left(SE(P_A)\right)^2 + \left(SE(P_C)\right)^2} \qquad (6)$$

SE(P_A - P_C) may be larger than 0.7; it is not capped at that value.

Example - Counts

Let's say we want to aggregate tracts 000200, 000400 and 001100 in Bronx County, NY (ST=36 & CTY=005), and look at table 1 line 26, the number of Blacks or African Americans.

TR	AVGEST	ACS_UNI	EST_C2K	C2K_UNI	DEFF00	AVGSE	ACS_UNI_SE
000200	620.67	3029.33	970	3590	2.2	215.10	473.98
000400	757.33	3515.67	900	3250	2.2	243.49	536.10
001100	50.00	104.33	80	400	1.8	66.28	73.56
Total	1428.00	6649.33	1950	7240			

$$P_A = \frac{1428.00}{6649.33} = 0.2148$$

$$P_C = \frac{1950}{7240} = 0.2693$$

$$P_A - P_C = 0.2148 - 0.2693 = -0.0545$$

The three tracts have different DEFF00 values, so we'll use the largest one, 2.2.

$$SE(P_C) = 2.2 \times \sqrt{\frac{5}{7240} \times 0.2693 \times (1 - 0.2693)} = 0.0256$$

$$SE(Num_A) = \sqrt{215.10^2 + 243.49^2 + 66.28^2} = 331.58$$

$$SE(Den_A) = \sqrt{473.98^2 + 536.10^2 + 73.56^2} = 719.35$$

$$SE(P_A) = \frac{1}{6649.33}\sqrt{(331.58)^2 - (0.2148)^2 \times (719.35)^2} = 0.0441$$

$$SE(P_A - P_C) = \sqrt{0.0441^2 + 0.0256^2} = 0.0510$$

Approximating the Standard Error - Totals

Some characteristics, such as the total number of persons or housing units, are not converted to proportions to compute the change. [Lines with METHOD = 'T'.]

Simply,

$$Est_A = \sum_{i=1}^{n} AVGEST_i$$

$$Est_C = \sum_{i=1}^{n} EST\_C2K_i$$

$$SE(Est_A) = \sqrt{\sum_{i=1}^{n} AVGSE_i^2}$$

$$SE(Est_C) = \sqrt{\sum_{i=1}^{n} NSE00\_C2K_i^2}$$

$$SE(Est_A - Est_C) \doteq \sqrt{\left(SE(Est_A)\right)^2 + \left(SE(Est_C)\right)^2}$$

Approximating the Standard Error – Ratios

Unfortunately, standard errors for most ratio lines (including means and percents) can't be calculated for aggregations of tracts given the information in the output files. However, standard errors for six of the ratio lines can be computed, using several different methods.

Group 1: lines 1-73 (average household size) and 3-6 (percent unemployed).

The standard errors of these ratios can be computed because both the numerator and denominator are separate lines in the profile. For line 1-73, the numerator is line 1-1 and the denominator is line 1-63. For line 3-6, the numerator is line 3-5 and the denominator is line 3-3.

$Num_A(i)$ = AVGEST from the numerator line
$Den_A(i)$ = AVGEST from the denominator line
$Num_C(i)$ = EST_C2K from the numerator line
$Den_C(i)$ = EST_C2K from the denominator line

$$SE(Num_A) = \sqrt{\sum_{i=1}^{n} \left(AVGSE_{num\ line}(i)\right)^2}$$

$$SE(Den_A) = \sqrt{\sum_{i=1}^{n} \left(AVGSE_{den\ line}(i)\right)^2}$$

$$SE(P_A) = \frac{1}{Den_A}\sqrt{\left(SE(Num_A)\right)^2 + \left(P_A\right)^2 \times \left(SE(Den_A)\right)^2}$$

Note the use of "+" instead of "-" in the formula for $SE(P_A)$. If $P_A=1$, then use $P_A=0$ when computing $SE(P_A)$.

For line 1-73:

$$SE(Num_C) = \sqrt{\sum_{i=1}^{n} \left(NSE00\_C2K_{num\ line}(i)\right)^2}$$

$$SE(Den_C) = \sqrt{\sum_{i=1}^{n} \left(NSE00\_C2K_{den\ line}(i)\right)^2}$$

$$SE(P_C) = \frac{1}{Den_C}\sqrt{\left(SE(Num_C)\right)^2 + \left(P_C\right)^2 \times \left(SE(Den_C)\right)^2}$$

For line 3-6:

$$SE(P_C) = DEFF00 \times \sqrt{\frac{5}{Den_C} \times P_C \times (1 - P_C)}$$

If $P_C < 0.02$ then use $P_C = 0.02$, and if $P_C > 0.98$ then use $P_C = 0.98$. If DEFF00 differs among the tracts being combined, use the largest value of DEFF00.

$$SE(P_A - P_C) \doteq \sqrt{\left(SE(P_A)\right)^2 + \left(SE(P_C)\right)^2}$$

Do *not* cap SE(PA) or SE(PC) at 0.7.

Group 2: lines 2-15 (percent high school graduate or higher), 2-16 (percent bachelor's degree or higher), 2-33 (percent employed, 21-64, with a disability), and 2-35 (percent employed, 21-64, with no disability).

For these four percents, the denominator is a separate line in the profile: 2-7, 2-7, 2-32 and 2-34, respectively.

$Den_A(i)$ = AVGEST from the denominator line
$SE(Den_A(i))$ = AVGSE from the denominator line
$P_A(i)$ = AVGEST/100 from the percent line
$Num_A(i) = Den_A(i) \times P_A(i)$
$SE(P_A(i))$ = AVGSE/100 from the percent line
$Den_C(i)$ = EST_C2K from the denominator line
$P_C(i)$ = EST_C2K/100 from the percent line
$Num_C(i) = Den_C(i) \times P_C(i)$

$$SE(Num_A(i)) = \sqrt{\left(Den_A(i) \times SE(P_A(i))\right)^2 - \left(P_A(i) \times SE(Den_A(i))\right)^2}$$

If $Den_A(i) = 0$, then set $SE(Num_A(i)) = SE(Den_A(i))$. We use a special approximation for the standard error when an estimate is zero. If the denominator is zero, the numerator is zero also, and their standard errors are the same. The above equation will not work in this case.

There is a chance that the value under the square root sign in the equation above will be negative. This could be due to rounding (it would not be negative if unrounded values were used) or other reasons. If that is the case for a tract, then $SE(P_A – P_C)$ can't be computed for the grouping that contains that tract.

$$SE(Num_A) = \sqrt{\sum_{i=1}^{n} \left(SE(Num_A(i)\right)^2}$$

$$SE(Den_A) = \sqrt{\sum_{i=1}^{n} \left(AVGSE_{den\ line}(i)\right)^2}$$

$$SE(P_A) = \frac{1}{Den_A}\sqrt{\left(SE(Num_A)\right)^2 + \left(P_A\right)^2 \times \left(SE(Den_A)\right)^2}$$

For the census estimate,

$$SE(P_C) = DEFF00 \times \sqrt{\frac{5}{Den_C} \times P_C \times (1 - P_C)}$$

If $P_C < 0.02$ then use $P_C = 0.02$, and if $P_C > 0.98$ then use $P_C = 0.98$. If DEFF00 differs among the tracts being combined, use the largest value of DEFF00.

$$SE(P_A - P_C) \doteq \sqrt{\left(SE(P_A)\right)^2 + \left(SE(P_C)\right)^2}$$

Do *not* cap $SE(P_A)$ or $SE(P_C)$ at 0.7. You may multiply $P_A – P_C$ and $SE(P_A – P_C)$ by 100% to convert them to percent estimates.

<u>Approximating the Standard Error – Medians</u>

Neither estimates nor standard errors of aggregations of tracts for median lines can be calculated with the data available.