

Appendix: Sampling Variability

The tables in this document present data that are from a 1 percent sample file drawn from the administrative records of the Social Security Administration.

Because of sampling variability, estimates based on sample data may differ from the figures that would have been obtained if all of the records, rather than specified samples, had been used. The standard error is a measure of sampling variability. About 68 percent of all possible probability samples selected with the same specifications will give estimates within one standard error of the figure obtained from the compilation of all records. Similarly, approximately 90 percent will give estimates within 1.645 standard errors, about 95 percent will give estimates within two standard errors, and about 99 percent will give estimates within 2.5 standard errors. The standard error of an estimate depends on the design elements, such as the method of sampling, sample size, and the estimation process.

Because of the large number of data cells tabulated from the sample files, it is not practical to calculate the standard error for every possible cell. However, standard errors for a large number of cells were estimated and used to fit regression curves to provide approximate standard errors, tabulated counts, and proportions.

The tables that follow provide a general order of magnitude of standard errors for similar estimates from the various sample files. Table A-1 presents approximate standard errors for the estimated number of persons from the 1 percent sample file. The reliability of an estimated percentage depends on the size of the percentage and on the size of the total on which the percentage is based. Table A-2 provides approximations of the standard errors of the estimated percentage of persons in the 1 percent sample file.

Table A-1.
Approximations of standard errors of estimated number of persons in the 1 percent file

| Size of estimate (inflated) | Standard error |
|-----------------------------|----------------|
| 500 | 250 |
| 1,000 | 300 |
| 2,500 | 500 |
| 5,000 | 800 |
| 7,500 | 900 |
| 10,000 | 1,100 |
| 25,000 | 1,700 |
| 50,000 | 2,400 |
| 75,000 | 3,000 |
| 100,000 | 3,400 |
| 250,000 | 5,400 |
| 500,000 | 7,800 |
| 750,000 | 9,600 |
| 1,000,000 | 11,100 |
| 5,000,000 | 25,800 |
| 10,000,000 | 36,900 |
| 25,000,000 | 57,700 |
| 50,000,000 | 76,100 |
| 75,000,000 | 82,900 |

Table A-2.
Approximations of standard errors of estimated percentage of persons in the 1 percent file

| Size of base (inflated) | 2 or 98 | 5 or 95 | 10 or 90 | 25 or 75 | 50 |
|-------------------------|---------|---------|----------|----------|------|
| 1,000 | 4.7 | 7.3 | 10.1 | 14.5 | 16.8 |
| 10,000 | 1.5 | 2.3 | 3.2 | 4.6 | 5.3 |
| 50,000 | 0.7 | 1.0 | 1.4 | 2.1 | 2.4 |
| 100,000 | 0.5 | 0.7 | 1.0 | 1.5 | 1.7 |
| 500,000 | 0.2 | 0.3 | 0.4 | 0.7 | 0.8 |
| 1,000,000 | 0.1 | 0.2 | 0.3 | 0.5 | 0.5 |
| 5,000,000 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 |
| 10,000,000 | (L) | 0.1 | 0.1 | 0.2 | 0.2 |
| 50,000,000 | (L) | (L) | (L) | 0.1 | 0.1 |
| 100,000,000 | (L) | (L) | (L) | (L) | (L) |

NOTE: (L) = less than 0.05 percent.