The 2006 Earnings Public-Use Microdata File

Data Dictionary and Field Descriptors

Part 1. Introduction

The 2006 Earnings Public-Use File (EPUF) is a systematic 1 percent random sample of all Social Security numbers issued prior to January 1, 2007. With a few minor exceptions, all of the values for the data fields in this file are from the Summary Segment of SSA's Master Earnings File, the administrative file used to determine an individual's eligibility status under the Social Security program and the amount of benefits paid out.

The EPUF consists of two separate, linkable sub-files—one with demographic and aggregate earnings information (demographic sub-file) and one with annual earnings information from 1951 to 2006 (annual earnings sub-file.) Each record on these sub-files has a unique, randomly assigned identifier allowing linking across both sub-files. The demographic sub-file contains 4,384,254 records, one for each individual included in EPUF. The annual earnings sub-file contains 60,326,474 earnings records with positive earnings values for the 3,131,424 individuals who had positive earnings for at least one year during 1951 to 2006. Years with zero earnings do not generate a record in the annual earnings sub-file.

All of the monetary values in the file have been bottom-coded, top-coded, or random rounded for data disclosure purposes. The rounding base used in the random rounding process depends on the amount of earnings.[1] The random rounding process provides some uncertainty about the actual values on the individual's earnings records and the interval of uncertainty increases as earnings increase.[2] All of the values for annual earnings are top-coded at the taxable maximum in a given year.

Several steps were taken to ensure that the random rounding process did not alter an individual's status in terms of:

1. worker (in covered employment) versus non-worker in a given year,
2. earnings below or at least equal to the taxable maximum in a given year,
3. whether or not the individual had any earnings in the 1937 to 1950 time period

All annual earnings values less than $100 are bottom-coded so rounding does not result in $0 earnings. Bottom coding means all earnings values less than $100 in a given year are replaced with the average value for all records with earnings less than $100 in that year. We also apply bottom coding to the aggregate values of Social Security Taxable Earnings from 1937 to 1950 that are less than $100.

To ensure that earnings below the taxable maximum are not random rounded up to the taxable maximum, all of the earnings values within the random rounding base ($100 or $1,000 depending on the year) of the taxable maximum are assigned the average value of all earnings records within that interval. For example, if the taxable maximum is $96,000 and an individual has earnings of $95,137.00, we assigned the average value for all earnings between $95,000 and $96,000 to the individual's record so it is not random rounded up to $96,000. We apply the same process to all other earnings records between $95,000 and $96,000 for that year.

[1] The dollar amounts for the random rounding base in EPUF are similar to those found in Zayatz, Laura. 2007. "Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update." *Journal of Official Statistics*, 23(2), 253-265. The rounding base for earnings between $1 and $999 in EPUF is higher than the base used in the article.

[2] Random rounding provides uncertainty when the amount of earnings reported to SSA is known. Thus, if someone has the actual value reported to SSA and traditional rounding is used, he/she will know the rounded value on EPUF. However, when random rounding is used, there are two possible values, only one of which is used in EPUF.

In-depth information about this file is available in:

The 2006 Earnings Public-Use Micro Data File: An Introduction, by Michael Compson

Part 2. Demographic and Aggregate Earnings Information

ID Identification Number

This is the first field in both sub-files. The ID data field allows data linking across both sub-files. We randomly assigned the Identification Number to each individual included in EPUF using a random number generator routine.

YOB Year of birth

YOB: Ranges from 1870 to 2006

SEX Sex of beneficiary

1:	Male
2:	Female
3:	Unspecified gender code

TC3750 Total Credits earned from 1937 to 1950

This field indicates the total number of Social Security credits earned by the individual based on his/her earnings from 1937 to 1950. The Social Security Administration estimates this field. We use credits and quarters of coverage interchangeably.

TC3750: Ranges from 0 to 56

TC5152 Total Credits earned from 1951 to 1952

This field indicates the total number of Social Security credits earned by the individual based on his/her earnings in 1951 and 1952. Annual amounts for the credits earned for each of these two years are not available in electronic format on the Master Earnings File. The Social Security Administration derives this field from administrative data.

TC5152: Ranges from 0 to 8

AE3750 Aggregate Social Security Taxable Earnings from 1937 through 1950

This field is the aggregate amount of Social Security Taxable Earnings from 1937 through 1950. SSA does not have annual values for taxable earnings during this time period available in electronic format on the Master Earnings File.

We made one of three potential adjustments to each value of this data field:

1. The value of all records whose aggregate taxable earnings from 1937 to 1950 was greater than $37,000 is top-coded and set equal to $41,500 (the rounded mean of all values greater than $37,000). Approximately one-half of 1 percent of all the values for this field is greater than the top-coded value.

2. The value of all records whose aggregate taxable earnings from 1937 to 1950 was greater than zero and less than $100 is bottom-coded and set equal to $39 (the rounded mean of all values less than $100).

3. The value of all remaining records is random rounded to multiples of $25 or $100 depending on the amount of taxable earnings.

    a. Aggregate taxable earnings greater than $100 and less than $1,000 are random rounded to a base of $25,
    b. Aggregate taxable earnings greater than $1,000 and less than $37,000 are random rounded to a base of $100.

Part 3. Annual Earnings Information

ID Identification Number

This is the first field in both sub-files. The ID data field allows data linking across both sub-files. We randomly assigned the Identification Number to each individual included in EPUF using a random number generator routine. For years in which an individual did not have earnings, he/she will not appear in the annual earnings sub-file. If an individual did not have *any* annual taxable earnings between 1951 and 2006 then his/her ID number is not in the annual earnings sub-file. There are 60,326,474 earnings records in this sub-file for the 3,139,001 individuals in the EPUF that have at least one year of annual taxable earnings from 1951 to 2006.

YEAR

The year when the individual had positive taxable earnings from 1951 to 2006.

YEAR: Ranges from 1951 to 2006

ANNUAL_QTRS

The annual number of quarters of coverage or credits earned from 1951 to 2006

This field indicates the total number of Social Security credits earned by the individual based on his/her earnings for a given year. The possible maximum value for any given year is four credits.

The annual estimates for the quarters of coverage earned in 1951 or 1952 are not available electronically on the Master Earnings File. Consequently, SSA estimates the values for the quarters of coverage earned in 1951 and 1952 and includes them in the demographic sub-file. In the annual earning sub-file, the annual values for quarters of coverage in 1951 and 1952 are set to a missing value.

ANNUAL_QTRS: Ranges from missing value, 0 to 4

ANNUAL_EARNINGS

Annual Social Security Taxable Earnings for each year from 1951 through 2006

This field contains only the positive values for annual Social Security Taxable Earnings up to the taxable maximum in a given year from 1951 to 2006.

ANNUAL_EARNINGS Ranges from $46 to $94,200

We made one of four potential adjustments to each value of this data field:

1. The value of all records whose annual taxable earnings were greater than the taxable maximum in a given year is top-coded at the taxable maximum.

2. Records with an annual value less than $100 are bottom-coded to the rounded mean of all values less than $100 for the given year.

3. With one exception - see number 4 below - records with annual taxable earnings greater than $100 and less than the taxable maximum for a given year are random rounded to multiples of $25, $100, $1,000 depending on the amount of taxable earnings.

4. Records with annual taxable earnings within the random rounding base of the taxable maximum for the given year are top-coded to the average value of all earnings records for that year between the taxable maximum and the taxable maximum minus the rounding base. For example, if the taxable maximum is $96,000 and an individual has annual taxable earnings equal to $95,250 his or her earnings would be set equal to the average value of all earnings records between $95,000 and $96,000 in that year.