Introduction to the Sample Files of the General-Purpose Sampling Database for Profit-Seeking Enterprise Income Tax Declaration

1. Sampling Dataset(One set of sampling data will be provided for each case, using the same sampling method)

Dataset Code	Dataset	Sample Dataset Name	Year
PRCT001	Profit-Seeking Enterprise Income Tax Profit and Loss Declaration Random Sampling Dataset	PRCT001_R_Year	095-104
PRCT002	Profit-Seeking Enterprise Income Tax Balance Sheet Random Sampling Dataset	PRCT002_R_Year	095-104
PRCT001	Profit-Seeking Enterprise Income Tax Profit and Loss Declaration Linked Panel Dataset	PRCT001_S_Year	095-104
PRCT002	Profit-Seeking Enterprise Income Tax Balance Sheet Linked Panel Dataset	PRCT002_S_Year	095-104

2. Sampling Design(For details, please refer to the sampling methods and validation plan in the General-Purpose Sampling Database for Profit-Seeking Enterprise Income Tax Declaration)

(1) Data Description:

i. Profit-seeking enterprise income tax is declared based on the "profit-seeking enterprise corporation," with data including total business revenue, net business revenue, operating profit, non-operating income, non-operating losses, profit and loss, taxable income, and tax calculation.
ii. Data from small-scale profit-seeking enterprises and organizations that do not meet the taxation standard is not included.
iii. The sampling method can be divided into:
    - **Random Sampling Dataset(R)**: Using the current year's profit-seeking enterprise income tax profit and loss declaration file as the population, random sampling is conducted each year according to sampling rules.

- **Linked Panel Dataset(S)**: The sampling basis is the profit-seeking enterprise income tax profit and loss declaration file for the starting year (Year 95), providing data traced back to the termination year (Year 104) based on the unified business number of the sampled enterprises and withholding units.

iv. The profit-seeking enterprise income tax balance sheet file is linked to the sampling results of the profit and loss declaration file for that year to obtain related data such as current assets, non-current assets, current liabilities, and non-current liabilities of the sampled enterprises and withholding units.

v. The amount columns in the profit-seeking enterprise income tax profit and loss declaration file are divided into "Book Amount" and "Self-Adjusted Amount." The "Book Amount" refers to the amounts recorded after the general ledger of the profit-seeking enterprise is settled, based on Generally Accepted Accounting Principles (GAAP). The "Self-Adjusted Amount" is based on the settled book amount, which is adjusted according to the Income Tax Act and related regulations, used for calculating amounts.

vi. For standard descriptions of tax industry codes, please refer to the Ministry of Finance's Statistical Office Tax Industry Standard Classification Query System. The website is: http://web02.mof.gov.tw/std/main.htm.

(2) Sampling Design and Validation:

i. **Random Sampling Dataset**

(1)Stratified Simple Random Sampling at a 95% Confidence Level: Based on the "Self-Adjusted Amount of Total Business Revenue," the population is divided, with the top 0.2% classified as the high-revenue group and the remaining 99.8% as the general revenue group.

- The high-revenue group consists of the top 0.2% sorted by self-adjusted total business revenue, divided into layers with each 10 records forming a layer, and 20% of the highest layers (layers n and n-1) are sampled to represent the median of that layer.
- For the general revenue group, stratified simple random sampling is conducted based on the tax industry standard.
- Samples from both the high-revenue and general revenue groups are combined to form the sample for the current year's profit-seeking enterprise income tax profit and loss declaration random file.

(2)Descriptive statistics, cumulative probability distribution, and K-S tests are used to verify the representativeness of the sampled data.

ii. **Linked Panel Dataset**

- Using the starting year's file (Year 95) as the sampling basis, similar slicing methods are applied to divide into high-revenue and general revenue groups for stratified random sampling to establish the sampling database for year t.
  - High-revenue group: Retain 75% of year t sample data for year t+1, removing enterprises that no longer exist, and using simple random sampling from the year t+1 population data to supplement until the sample count matches year t. This process continues for subsequent years.
  - General revenue group: Similar to the high-revenue group, but stratified random sampling is applied.
  - Samples from both groups are merged to form the sample for the current year's profit-seeking enterprise income tax linked panel dataset.
- Descriptive statistics, chi-square test, and T-test are used to verify the representativeness of the sampled data.

iii. **Sample Size for Sampling Files**:

- Random Sampling Dataset: One set of 20% of the profit and loss declaration file for each application case will be provided.
- Linked Panel Dataset: One set of 20% of the profit and loss declaration file based on the starting year (Year 95) will be provided for each application case.

3. Available Fields

Provides commonly used fields for research (see the database user manual for details).

4. Payment Method:

(1) The data processing fee, monitoring room equipment usage, and management fees are charged according to the "Ministry of Finance's Financial Information Center Tax Information Fee Standard."

(2) The data usage period is one year, limited to the actual approval period for each case.