ATTACHMENT A

# Probability Sample Documentation Standards

## Sampling Plan

A written sampling plan should be prepared and formalized prior to the execution of the sample. A plan would include the following:

- The objective of the plan including a description of what value is being estimated and for which tax year(s) the estimate is applicable.
- Population definition and reconciliation of the population to the tax return.
- Definition of the sampling frame.
- Definition of the sampling unit.
- Source of the random numbers, the starting point or seed, and the method used in selecting them.
- Sample size, along with supporting factors in the determination.
- Method used to associate random numbers to the frame.
- Steps to be taken to insure that the serialization of the frame is carried out independent of the drawing of random numbers.
- Steps to be taken in evaluating the sampling unit.
- The appraisal method(s) to be used in appraising the sample.

## Sample Execution Documentation

The execution of the sample must be documented and include information for each of the following:

- The seed or starting point of the random numbers.
- The pairing of random numbers to the frame along with supporting information to retrace the process.
- List of the sampling units selected and the results of the evaluation of each unit.
- Supporting documentation such as notes, invoices, purchase orders, project descriptions etc., which support the conclusion reached about each sample item.
- The calculation of the projected estimate(s) to the population, including the computation of the standard error of the estimate(s).
- A statement as to any slips or blemishes in the execution of the sampling procedure and any pertinent decision rules.
- Computation of all associated adjustments. (An example of an associated adjustment would be the amount of depreciation allowable based on a probability determination of an amount capitalized).

TECHNICAL APPENDIX

# FORMULAS

| **UNSTRATIFIED (SIMPLE RANDOM SAMPLE) MEAN ESTIMATOR** | **STRATIFIED MEAN ESTIMATOR** |
|---|---|

**Sample Mean of Audited Amounts**

$$\bar{x} = \frac{x_j}{n}$$

**Estimate of Total Audited Amount**

$$\hat{X}_M = N\bar{x}$$

$$\hat{X}_{MS} = (N_i \bar{x}_i)$$

**Estimated Standard Deviation of the Audited Amount**

$$S_X = \sqrt{\frac{[\ (x_j^2)] - n(\bar{x}^2)}{n-1}}$$

**Estimated Standard Error of the Total Audited Amount**

$$\hat{\sigma}(\hat{X}_M) = \frac{NS_x\sqrt{1 - n/N}}{\sqrt{n}}$$

$$\hat{\sigma}(\hat{X}_{MS}) = \sqrt{\sum\left[N_i(N_i - n_i)\frac{S_{x_i}^2}{n_i}\right]}$$

**Achieved Precision of the Total Audited Amount**

$$A'_M = \frac{NU_R S_x\sqrt{1 - n/N}}{\sqrt{n}}$$

$$A'_{MS} = U_R\sqrt{\sum\left[N_i(N_i - n_i)\frac{S_{x_i}^2}{n_i}\right]}$$

TECHNICAL APPENDIX

## FORMULAS

| **UNSTRATIFIED (SIMPLE RANDOM SAMPLE) DIFFERENCE ESTIMATOR** | **STRATIFIED DIFFERENCE ESTIMATOR** |
|---|---|

### Estimate of Total Difference

Unstratified:

$$\hat{D} = N\,\bar{d}$$

Stratified:

$$\hat{D}_S = \quad (N_i\,\overline{d_i}\,)$$

### Estimate of Total Audited Amount

Unstratified:

$$\hat{X}_D = Y - \hat{D}$$

Stratified:

$$\hat{X}_{DS} = Y - \hat{D}_s$$

### Estimated Standard Deviation of the Difference Amount

$$S_D = \sqrt{\frac{\left[\quad (d_j^2)\right] - n(\bar{d}^2)}{n-1}}$$

### Estimated Standard Error of the Difference Amount

Unstratified:

$$\hat{\sigma}(\hat{D}) = \frac{NS_D\sqrt{1 - n/N}}{\sqrt{n}}$$

Stratified:

$$\hat{\sigma}(\hat{D}_S) = \sqrt{\sum\left[N_i\,(N_i - n_i)\,\frac{S_{D_i}^2}{n_i}\right]}$$

### Achieved Precision of the Difference Amount

Unstratified:

$$A'_D = \frac{NU_R S_D\sqrt{1 - n/N}}{\sqrt{n}}$$

Stratified:

$$A'_{Ds} = U_R\sqrt{\sum\left[N_i\,(N_i - n_i)\,\frac{S_{D_i}^2}{n_i}\right]}$$

TECHNICAL APPENDIX

FORMULAS

| **UNSTRATIFIED (SIMPLE RANDOM SAMPLE) RATIO ESTIMATOR** | **STRATIFIED COMBINED RATIO ESTIMATOR** |
|---|---|

**Estimated Ratio of Audited Amount to Recorded Amount**

$$R = \frac{x_j}{y_j} = 1 + \frac{d_j}{y_j}$$

$$\hat{R}_C = \frac{(N_i \, \overline{x}_i)}{(N_i \, \overline{y}_i)} = 1 + \frac{(N_i \, \overline{d}_i)}{(N_i \, \overline{y}_i)}$$

**Estimate of Total Audited Amount**

$$\hat{X}_R = Y\hat{R}$$

$$\hat{X}_{RC} = Y\hat{R}_C$$

**Estimated Standard Deviation of the Ratio**

$$S_R = \sqrt{\frac{(x_j^2) + \hat{R}^2 \quad (y_j^2) - 2\hat{R} \quad (x_j y_j)}{n-1}}$$

**Estimated Standard Deviation of the Ratio in *i*th Stratum**

$$S_{RC_i} = \sqrt{\frac{[( \quad x_{ij}^2) - (x_{ij}^2 / n_i)] + [(\hat{R}_C^2 \quad y_{ij}^2) - (y_{ij}^2 / n_i)] - [(2\hat{R}_C \quad x_{ij} y_{ij}) - (x_{ij} * y_{ij} / n_i)]}{n_i - 1}}$$

**Estimated Standard Error of the Ratio Amounts**

$$\hat{\sigma}(\hat{X}_R) = \frac{NS_R \sqrt{1 - n/N}}{\sqrt{n}}$$

$$\hat{\sigma}(\hat{X}_{RC}) = \sqrt{\sum \left[ N_i \, (N_i - n_i) \frac{S_{RC_i}^2}{n_i} \right]}$$

**Achieved Precision of the Ratio Amounts**

$$A'_R = \frac{NU_R S_R \sqrt{1 - n/N}}{\sqrt{n}}$$

$$A'_{RC} = U_R \sqrt{\sum \left[ N_i \, (N_i - n_i) \frac{S_{RC_i}^2}{n_i} \right]}$$

# TECHNICAL APPENDIX
# FORMULAS

**UNSTRATIFIED (SIMPLE RANDOM SAMPLE) REGRESSION ESTIMATOR** | **STRATIFIED COMBINED REGRESSION ESTIMATOR**

---

**Estimated Regression Coefficient**

$$b = \frac{[\quad (x_j y_j)] - n\bar{x}\bar{y}}{[\quad (y_j^2)] - n(\bar{y}^2)} = 1 + \frac{[\quad (d_j y_j)] - n\bar{d}\bar{y}}{[\quad (y_j^2)] - n(\bar{y}^2)}$$

$$b_C = \frac{\left[\sum N_i (N_i - n_i) \frac{S_{XY_i}}{n_i}\right]}{\left[\sum N_i (N_i - n_i) \frac{S_{Y_i}^{2}}{n_i}\right]} = 1 + \frac{\left[\sum N_i (N_i - n_i) \frac{S_{DY_i}}{n_i}\right]}{\left[\sum N_i (N_i - n_i) \frac{S_{Y_i}^{2}}{n_i}\right]}$$

**Estimate of Total Audited Amount**

$$\hat{X}_G = N\bar{x} + b(Y - N\bar{y})$$

$$\hat{X}_{GC} = \quad (N_i \bar{x}_i) + b_C[Y - \quad (N_i \bar{y}_i)]$$

**Estimated Standard Deviation of the Regression Amounts**

$$S_G = \sqrt{\frac{1}{n-2}\left[[\Sigma(x_j^2)] - n(\bar{x}^2) - \frac{(\Sigma(x_j y_j) - n\bar{x}\bar{y})^2}{\Sigma(y_j^2) - n(\bar{y}^2)}\right]}$$

**Estimated Covariance between the Audited and Recorded Amounts in *i*th Stratum**

$$S_{XY_i} = \frac{[\quad (x_{ij} y_{ij})] - n_i \bar{x}_i \bar{y}_i}{n_i - 1}$$

**Estimated Standard Deviation between the Audited and Recorded Amounts in *i*th Stratum**

$$S_{GC_i} = \sqrt{S_{X_i}^{2} - 2b_C S_{XY_i} + b_C^{2} S_{Y_i}^{2}}$$

**Estimated Standard Error of the Audited and Recorded Amounts**

$$\hat{\sigma}(\hat{X}_G) = \frac{NS_G\sqrt{1 - n/N}}{\sqrt{n}}$$

$$\hat{\sigma}(\hat{X}_{GC}) = \sqrt{\sum\left[N_i (N_i - n_i) \frac{S_{GC_i}^{2}}{n_i}\right]}$$

**Achieved Precision of the Audited and Recorded Amounts**

$$A'_G = \frac{NU_R S_G\sqrt{1 - n/N}}{\sqrt{n}}$$

$$A'_{GC} = U_R\sqrt{\sum\left[N_i (N_i - n_i) \frac{S_{GC_i}^{2}}{n_i}\right]}$$

TECHNICAL APPENDIX

## Definition of Symbols

| TERM | DEFINITION |
|---|---|
| **n** | **Sample Size** |
| **N** | **Population Size** |
| **x** | **The value of the sampling unit that is being used as the primary variable of interest. In audit sampling, this would be the audited (or revised) value of the transaction. In LIFO Index samples, it is represented by the end of year value.** |
| **y** | **The value of the sampling unit that is being used as the "paired" variable that is related to the variable of interest. In audit sampling, this would be the reported (or original) value of the transaction. For LIFO Index samples, it is represented by the beginning of year value. This variable is used with the difference, ratio, and regression estimators.** |
| **d** | **The value of the sampling unit that is the difference between "paired" variable (y) and the variable of interest (x). That is,** $d = y - x$ **In audit sampling, this would be the difference (or the change) of each transaction's value.** |
| **X** | **The total value of the primary variable of interest. In audit sampling, this would be the estimated total audited value of the population. Typically, this value is not known for the entire population and is estimated based on the probability sample selected.** |
| **Y** | **The total value of the variable that is paired with variable of interest. In audit sampling, this would be the total reported value of the population. Typically, this value is known for the entire population and may be estimated based on the probability sample selected.** |
| **D** | **The total value of the difference between the "paired" variable and the variable of interest. In audit sampling, this would be the estimated total difference of the population. Typically, this value is not known for the entire population and is estimated based on the probability sample selected.** |
| $U_R$ | **The confidence coefficient which is based on either the Student's *t*-distribution or the normal distribution. For example, a 95% one-sided confidence coefficient based on the normal distribution is 1.645. This term is often referred to as the *t*-value and the *z*-value.** |