



## **SAMPLING GUIDELINES FOR RESEARCH EXPENSE CREDIT CASES**

### **PURPOSE**

This document provides taxpayers with guidance regarding the use and evaluation of statistical samples and sampling estimates for research expense credits. If statistical sampling is determined to be appropriate, a taxpayer may use only the statistical sampling procedures set forth in this document.

### **SCOPE**

Taxpayers may use statistical sampling procedures to support items on its IT-20REC tax return or to calculate the research expense credit for a Claim for Refund. This procedure is limited to variable sampling and is effective for taxable years ending on or after January 1, 2019. The Department will use the criteria set forth in this document in determining whether to accept a statistical sampling estimate as adequate substantiation for a return or Claim for Refund position. Statistical samples that fail to meet these criteria will be rejected.

### **APPLICATION**

When permitted by the Department, a taxpayer may use statistical sampling in establishing, with respect to its research expense credit, by following the procedures provided in Appendix A, and Appendix B.

In order to conduct a statistical sample the research credit must involve more than 100 sampling units. The appropriateness of using a statistical sample, to either support or be primary evidence of an amount, is a facts and circumstances determination. Factors to be used in determining whether a statistical sample is appropriate include, but are not limited to, the following: the time required to analyze large volumes of data; the cost of analyzing data; and the other books and records that may independently exist or have greater probative value. Statistical samples generally will not be considered appropriate if evidence is readily available from another source that can be demonstrated to provide a more accurate answer, or if the use of sampling does not conform to applicable financial accounting standards (e.g., Generally Accepted Accounting Principles (GAAP)).

The sampling unit must be determined to conduct a statistical sample. Acceptable sampling units are:

- (1) Indiana Projects
- (2) Wages

- (3) Contract Research Expenses
- (4) Supplies

The statistical sample should include a minimum of 30 records in each. This requirement does not apply to strata tested on a 100% basis (i.e., the entire population of a stratum is selected for evaluation).

Taxpayers must determine whether the final estimate represents a valid estimate. In general, a final estimate will be considered valid (without regard to audit adjustment(s)) provided that all of the following conditions are met.

**Taxpayers must maintain all of the proper documentation to support the statistical application, sample unit findings, and all aspects of the sample plan. Proper supportive documentation generally includes all of the information contained in Appendix A and Appendix B.**

The estimate must be based on a statistical sample, in which each sampling unit in the population has a known (non-zero) chance of selection, using either a simple random sampling method or stratified random sampling method.

The estimate must be computed at the least advantageous 95% one-sided confidence limit. The “least advantageous” confidence limit is either the upper or lower limit that results in the least benefit to the taxpayer. If the relative precision for a sampling plan does not exceed 10%, the point estimate may be used in place of the least advantageous 95% one-sided confidence limit. When the relative precision is less than 15% and greater than 10%, the estimate will be computed as an amount between the least advantageous 95% one-sided confidence limit and the point estimate determined as follows:

$$\text{Estimate} = \text{Point Estimate} - (\text{Relative Precision} - .10) / .05 * (\text{Point Estimate} - \text{Least Advantageous 95\% One-Sided Confidence Limit})$$

Although many methods exist to estimate population values from the sample data, only the following estimators will be considered for acceptance:

- (i) Variable estimators permitted include the Mean (also known as the direct projection method);
- (ii) Difference (using “paired variables”);
- (iii) (combined) Ratio (using a variable of interest and a “correlated” variable); and
- (iv) (combined) Regression (using a variable of interest and a “correlated” variable).

The first variable used for the difference, ratio and regression estimators must be the variable used in the mean estimator. The second variable used for the difference,

ratio, and regression estimators must be a variable that can be paired with the first variable and should be related to the first variable.

To be accepted by the Department as a method to estimate population values from the sample data, taxpayers who choose to use methods (iii) or (iv) described above must first demonstrate that the statistical bias inherent in those methods is negligible.

Use estimate with smallest overall standard of error. Of all the final estimates determined as qualifying, the estimate with the smallest overall standard of error, as an absolute value, will generally be used (i.e., the size of the estimate is irrelevant in the determination of the value to be reported). Situations may exist when only a single estimator may be appropriate for the plan objective. In those specialized situations, the relevant estimator may be evaluated without consideration of other methods.

Confidence limits are calculated by adding and subtracting the precision of the estimate from the point estimate when precision is determined by multiplying the standard error by (i) the 95% one-sided confidence coefficient based on the Student's t-distribution with the appropriate degrees of freedom, or (ii) 1.645 (i.e., the normal distribution), assuming the sample size is at least 100 in each non-100% stratum.

For either the (combined) Ratio or Regression methods, in order to demonstrate that little statistical bias exists, the following applies after excluding all strata tested on a 100% basis:

- (i) the total sample size of all strata must be at least 100 units;
- (ii) each stratum for which a population estimate is made should contain at least 30 sample units;
- (iii) the coefficient of variation of the paired variable must be 15% or less;
- (iv) the coefficient of variation of the primary variable of interest, represented by either the corrected value or the difference between the reported and corrected values in common accounting situations must be 15% or less; and
- (v) for only the (combined) Ratio method, the reported values of the units must be of the same sign. Definitional formulas for the paired variable (described in (iii) above), and the corrected value and the difference between the reported and corrected values (each described in (iv) above).

The relative precision for each estimator is commonly calculated by dividing the precision at the 95% one-sided confidence limit (sometimes referred to as sampling error) of the estimate by the estimator. When an estimate may be calculated using either a corrected value or difference perspective, as in the case of Ratio and Regression methods, or solely a corrected value perspective as in the case of a Mean

method, the test will be applied on the basis of a difference perspective. In these cases the numerator of the calculation is the sampling error of the adjustment and the denominator the point estimate of the adjustment.

For the purpose of the 10% relative precision test, any stratum, when the sampling units or the process of evaluating the sampling units are different from those in other strata, must be excluded in calculating the relative precision.

When using simple random samples, the confidence limits are determined using the Hyper-geometric distribution. For stratified random samples, when at least two strata are sampled (i.e., not 100% samples), the confidence limits must be determined using the normal distribution approximation.

For the normal distribution approximation, the precision is calculated by multiplying the standard error by (i) the 95% one-sided confidence coefficient based on the Student's t-distribution with the appropriate degrees of freedom, or (ii) 1.645 (i.e., the normal distribution), assuming the sample size is at least 100 in each non-100% stratum.

The Department's allowance of a taxpayer's estimate does not correspondingly require acceptance by the Department of the taxpayer's use of an estimate for the determination of associated adjustments, allocation, or subdivision of the findings for other purposes.

This document only addresses the statistical requirements that must be met for a statistical sample to meet preliminary acceptance. It is not intended to further require acceptance of individual sample unit determinations. Valuation determinations remain subject to independent verification along with other non-statistical issues such as missing sampling items. Likewise, the statistical procedures followed may be examined and adjusted if the procedures are found to be in error. Any fatal error in statistical methodology that renders the statistical sample invalid will preclude the use of any statistical estimate based on the sample and will only allow for consideration of the sample findings on an actual basis. When a statistical sample is determined to be not appropriate and is raised as an issue, the examining agent may pursue a more accurate determination or allow the findings of units examined on an actual basis. The computational validity of the estimator should still be considered and addressed along with other alternative issues in un-agreed cases.

This document does not preclude the Department from raising or pursuing any income, employment, or other tax issues identified in the review of a statistical sample. It is recognized that existing industry practices and specific taxpayers may be using techniques that are not covered by this document. If a taxpayer employed a statistical sample or method not covered by this document, then the estimate may be referred to audit division for resolution or issue development.

## APPENDIX A

Taxpayers are required to have a written sampling plan prior to the execution of a sample. The plan must include the following items:

- (1) The objective of the plan including a description of what value is being estimated and for which tax year(s) the estimate is applicable;
- (2) Population definition and reconciliation of the population to the tax return;
- (3) Definition of the sampling frame;
- (4) Definition of the sampling unit;
- (5) Source of the random numbers, the starting point or seed, and the method used in selecting them. **Seed number used must be the 10-digit business phone number of the individual preparing the sample. The population must be sorted in ascending order by the project names included in the sampling frame. Random number generator must be Mersenne Twister algorithm;**
- (6) Sample size, along with supporting factors in the determination;
- (7) Method used to associate random numbers to the frame;
- (8) Steps to be taken to insure that the serialization of the frame is carried out independent of the drawing of random numbers;
- (9) Steps to be taken in evaluating the sampling unit; and
- (10) The appraisal method(s) to be used in appraising the sample.

## APPENDIX B

Taxpayers must retain adequate documentation to support the statistical application, sample unit findings, and all aspects of the sample plan and execution. The execution of the sample must be documented and include information for each of the following:

- (1) The seed or starting point of the random numbers;
- (2) The pairing of random numbers to the frame along with supporting information to retrace the process;
- (3) List of the sampling units selected and the results of the evaluation of each unit;
- (4) Supporting documentation such as notes, invoices, purchase orders, project descriptions, etc., which support the conclusion reached about each sample item;
- (5) The calculation of the projected estimate(s) to the population, including the computation of the standard error of the estimate(s);
- (6) A statement as to any slips or blemishes in the execution of the sampling procedure and any pertinent decision rules; and
- (7) Computation of all associated adjustments.