Florida Department of Revenue

Auditing in an Electronic Environment (e-Auditing)

and Stratified Statistical Sampling

Purpose

The purpose of this booklet is to inform taxpayers, practitioners, and other interested parties of the sampling methods that can be used by the Department and to provide a general understanding of what to expect if the auditor uses electronic auditing and stratified statistical sampling. Although the emphasis is upon sales and use tax, the same procedures may be used for audits of other taxes administered by the Florida Department of Revenue.

Credits

Designed and developed by the Florida Department of Revenue, General Tax Administration Statistics Team.

Sponsored by:

State of Florida
Department of Revenue
General Tax Administration Program
Compliance Review Process
Tallahassee, Florida

Florida Department of Revenue
Auditing in an Electronic Environment (e-Auditing)
And Stratified Statistical Sampling

Table of Contents

Electronic Auditing

(e-Auditing)

Electronic auditing (e-Auditing) is computer-assisted auditing, where electronic records are used to complete all or part of the audit. The following topics regarding electronic audits performed by the Department are explained in this booklet:

- Benefits of e-Auditing.
- Statutory authority.
- Confidentiality of data.
- Electronic audit candidates.
- Electronic audit techniques.

Benefits of e-Auditing

The decision of whether the Department will pursue an electronic audit instead of a paper document audit is based on the nature of the taxpayer's records. Electronic audits can reduce the collective effort by the taxpayer and the Department to complete the audit. Shorter audit time results in cost savings. In many instances, detailed audit procedures can be performed using electronic data in the same amount of time as it takes to audit a sample of paper documents.

Legal Authority

Any audit where the taxpayer's records are maintained electronically has the potential of being an electronic audit. If electronic data is available in any form, it may be used. Section 213.35, Florida Statutes, states:

> Each person required by law to perform any act in the administration of any tax enumerated in s. 72.011 shall keep suitable books and records relating to that tax, such as invoices, bills of lading, and other pertinent records and papers, and shall preserve such books and records until expiration of the time within which the department may make an assessment with respect to that tax pursuant to s. 95.091(3).

Part II of Rule Chapter 12-24, Florida Administrative Code (F.A.C.), provides guidance concerning a taxpayer's recordkeeping and retention responsibilities. This rule chapter describes requirements when "all or a part of the taxpayer's records are received, created, maintained or generated through computer, electronic, and imaging processes and systems." Rule 12-24.023, F.A.C., states in part:

> (2) If a taxpayer maintains records required to be retained under this chapter in both machine-sensible and hardcopy formats, the taxpayer shall make the records available to the Department in machine-sensible format upon request of the Department.

Rule 12-24.026, F.A.C., addresses a taxpayer's responsibilities concerning Department access to machine-sensible records. Access may be provided in several ways:

- The taxpayer can make equipment and personnel available so the auditor may access the records.
- The taxpayer can arrange with a third party to make equipment and personnel available so the auditor may access the records.
- The taxpayer can convert the records to a format specified by the auditor and provide the converted records on a medium agreed to by the auditor.
- The taxpayer and the auditor may agree on some other means of providing access.

Confidentiality of Data

The confidentiality of all taxpayer information, whether on paper or electronic medium, is ensured by s. 213.053, F.S. Any violation of confidentiality by any Department employee is a misdemeanor of the first degree and punishable under s. 775.082, F.S. or s. 775.083, F.S.

If the taxpayer has concerns about the confidentiality of electronic data when sending by common carrier or the U.S. Postal Service, data may be sent with no return address and a unique identifying number (available from the Department) on the CD or tape. In addition, the file layout may be sent separately from the data. This maintains anonymity in the unlikely event that data is lost by the common carrier or the U.S. Postal Service.

Electronic Audit Candidates

Taxpayers who maintain their records on a computer are candidates for an electronic audit. When a taxpayer's transactions are not completely electronic, an electronic audit of the available areas, such as purchases, can be performed. After the taxpayer completes an *Electronic Audit Survey,* the Department can determine whether an electronic audit may be conducted. **It is important to note that the electronic auditing techniques described in this booklet can only be used when taxpayer records have been converted to a format specified by the Department and provided to the Department on an agreed upon medium.**

Acceptable Data Format and Supported Media

If electronic data is located in a mainframe computer system, a flat file (non-relational) is requested. The format of the files can be EBCDIC, ASCII, Comma Delimited, Tab Delimited, or Space Delimited. Acceptable data format and media are shown in Table 1.

Table 1

Data Format	Supported Media
• Mainframe data EBCDIC, ASCII (Comma delimited, tab delimited, or space delimited) • Text file • Print file • Spreadsheet files (Excel, Quattro-Pro, Lotus, etc.) • Database files (dBase, Access, etc.)	• 3480/3490 tape cartridges • 9 track • 4 mm DAT (4GB capacity) • Compact Disc • 1 or 2 gig JAZ disk • 100 meg ZIP disk

Completion of an electronic file layout form is required for conversion of flat files. If flat files are not available, but the taxpayer can provide a printout of the needed information, a print file is requested. The Department can convert print files to database files. If neither flat files nor print files are available, the required electronic records are requested in a spreadsheet or database format (i.e., Excel, Access, etc.).

Data Integrity Tests

The Department tests the taxpayer's electronic data for accuracy. Tests may include the following:

- Vouching and reconciling a random number of electronic transactions with the taxpayer's source documents.

- Tracing and reconciling a random number of the taxpayer's source documents to the electronic data.

- Comparing summary totals of the electronic data to general ledger balances, tax returns, or other appropriate financial reports. The dollar amounts do not always match up perfectly. In most cases, the differences are simply caused by accounting timing issues, manual adjustments, etc. Any differences between the totals of the electronic data and the general ledgers or reports are discussed.

Electronic Audit Techniques

The tax auditor selects one or a combination of the following electronic audit techniques, based on the taxpayer's records:

- Electronic detail audit.
- Stratified statistical sample.
- Electronic judgmental sample.

It is common to use more than one approach in the same audit. For example, sales and fixed assets could be audited using the electronic detail approach, while expenses could be audited using stratified statistical sampling.

Sales Data

The following fields should be included in the electronic data. These fields, or some variation, are normally used to conduct an electronic audit of sales transactions:

- List of resale certificates on file.
- Customer name.
- Invoice date.
- Invoice number.
- Any identifying number needed by the taxpayer to locate the source documents.

- Ship to name.
- Ship to address.
- Ship to city.
- Ship to state.
- Ship to zip code.
- Invoice line item information including:
  - Line item number.
  - Product number.
  - Product description.
  - Quantity.
  - Taxable status of product.
  - Amount.
  - Tax collected (either in total or split by state and county, if available).

An electronic detail audit is generally the preferred technique for sales data. When performing an electronic audit, it may be possible to review every transaction in a file, precluding the need for sampling. If a detail audit can be performed within approximately the same amount of time as a sample, a detail audit should be performed whenever possible. The following are examples of detail audit procedures that can be performed on sales transactions:

- Compare total tax invoiced to the sales tax returns and general ledger.
- Recalculate total tax due and compare with the total tax invoiced.
- Export all taxed sales transactions to a separate file to review for applicable county tax rates.
- Export exempt sales transactions to a separate file to examine for exempt products and/or resale certificates.
- Create a unique index of customers to compare with the resale certificate file.
- Create a unique index of products to verify taxable and exempt products.

When auditing sales for resale or exemption certificates, a list of customers may be obtained to verify the certificates. When certificates are available electronically, they may be matched to the customer list and the exceptions examined in detail. Additionally, the customers that match the certificate list can be sampled to verify the validity of the certificates on the list.

After identifying transactions that will be included in the audit workpapers as exceptions, an exhibit will be set up in the Department's audit software program and the appropriate transaction detail imported from the data, eliminating the need for the auditor to manually enter the transactions. The software will compute additional tax due, including applicable penalty and interest on the transactions.

If the sales data is too voluminous to be efficiently audited using detail audit techniques, stratified statistical sampling can be used. Stratified statistical sampling allows for a detailed

examination of all large dollar invoices above a certain threshold amount, with the remaining small dollar invoices being randomly sampled.

Purchases Data

Purchases data may include fixed assets, expense purchases, and rental transactions. The following fields, or some variation, are normally used to conduct an electronic audit of purchases.

- Vendor name.
- Invoice date.
- Accounting journal entry date (or date paid).
- Vendor's invoice number.
- Any identifying number needed by the taxpayer to locate the source documents (voucher number, batch number, etc.).
- Division or cost center making purchase.
- Expense account.
- Description of item(s) purchased.
- Expensed amount.
- Total invoice and/or voucher amount.
- Sales tax paid or accrued information, if available.

Florida Statutes require a detailed examination of fixed assets. If fixed asset data is captured in specified general ledger account(s), the fixed asset transactions in these accounts are filtered from the remaining transactions and audited in detail. Since sales tax is a transaction tax, tax paid or accrued on each transaction must be verified. If sales or use tax is not recorded electronically on the fixed asset transactions, source documents may be provided so the auditor can verify proper tax treatment.

When auditing expense purchase records, stratified statistical sampling can generally be used. As with fixed assets, tax paid or accrued on expense purchases is generally not recorded separately in the electronic data. For this reason, source documents may also be necessary to verify proper tax treatment.

Real property rental transactions are generally separated into general ledger accounts that may be filtered from the data and audited in detail. Voluminous tangible personal property rental transactions may be included in the expense purchases population to be sampled.

Stratified Statistical Sampling

$$(a + b)^2 = a^2 + 2ab + b^2$$

Stratified statistical sampling is a methodology in which items in the population are classified into separate subgroups (strata) based on one or more important characteristics, such as dollar value. Taxpayers with adequate electronic records are candidates for stratified statistical sampling. If a detailed examination of the electronic records can be performed in approximately the same amount of time as a sample, a detail audit should be performed.

Benefits of Stratified Statistical Sampling

Statistical samples usually require smaller sample sizes than non-statistical judgmental samples. Smaller sample sizes normally mean less audit time and more efficient audits. Statistical sampling uses random selection of sample points where each record in the population has an equal probability of being selected; therefore, a random sample is theoretically representative of the population. Because extreme values are examined in detail, they are not projected.

Statistical sampling allows for the calculation of sample precision. A 95% confidence level is used by Department software to calculate sample size and precision. If all other parts of the sample-size formula are held constant, decreasing the confidence level to 90% or 85% results in only a negligible change in sample size.

Stratified Statistical Sampling Candidates

In order to employ stratified statistical sampling procedures as outlined here, the taxpayer's records must be adequate and available electronically. When using stratified statistical sampling, each item in a population must be stratified by dollar amount. Stratifying paper copies of records by dollar amount is not cost effective. Other sampling techniques can be employed for audits with adequate paper records.

When electronic data is obtained for only a portion of the audit period, stratified statistical sampling may be used for that period of time where electronic data is obtained and the results will be statistically valid for that period. If it is determined that the results of the sample are the best information available to project to the period where electronic data is not available, projecting the results of the stratified statistical sample to the period of time where electronic data is not available may be an acceptable non-statistical judgmental application. While this is an acceptable judgmental application, such an application would not be statistically valid.

In this situation, two separate exhibits are prepared to document the sample findings. One exhibit contains the stratified statistical sample results; the second exhibit contains the projection of the statistical sample results to the remaining audit period.

Stratified statistical sampling may be the most efficient method for completing a refund verification audit or an audit of taxpayers who have adopted an alternative reporting method of use tax. This alternative method involves applying a set taxable percentage to total expense purchases for the month and then applying an applicable tax rate to the result. Special authorization from the Department to use an alternate method of reporting use tax is not required. However, in accordance with s. 212.07(8), F.S., if a taxpayer uses an alternative reporting method then the taxpayer remains liable for the correct tax on all transactions.

When use tax is reported in this manner, the ability to verify the amount of use tax paid on each transaction is lost. However, stratified statistical sampling procedures may be used to project the amount of use tax that should have been paid. This projection should be compared to the use tax that was accrued and remitted to determine any deficiency or overpayment.

Statutory Basis for Sampling

Under Florida Statutes, any taxpayer who pays sales and use tax to the Department and files returns is required to keep records and other information which will allow the Department to determine the correct amount of tax due. Section 212.12(6)(c), F.S., provides that the Department may sample a taxpayer's records, except fixed assets, when the records are adequate but voluminous. When records are adequate but voluminous, both underpayments and overpayments found in the sample may then be projected over the entire audit period.

Per s. 212.12(6)(c)1, F.S., if the taxpayer's records are adequate but voluminous, and the taxpayer disagrees with the sampling method used in the audit, the taxpayer may request a review of the sampling method during the protest period.

Section 212.12(6)(c), F.S., states:

> 1. If the records of a dealer are adequate but voluminous in nature and substance, the department may sample such records, except for fixed assets, and project the audit findings derived therefrom over the entire audit period to determine the proportion that taxable retail sales bear to total retail sales or the proportion that taxable purchases bear to total purchases. In order to conduct such a sample, the department must first make a good faith effort to reach an agreement with the dealer, which agreement provides for the means and methods to be used in the sampling process. In the event that no agreement is reached, the dealer is entitled to a review by the executive director.
>
> 2. For the purposes of sampling pursuant to subparagraph 1., the department shall project any deficiencies and overpayments derived therefrom over the entire audit period. In determining the dealer's compliance, the department shall reduce any tax deficiency as derived from the sample by the amount of any overpayment derived from the sample. In the event the department determines from the sample results that the dealer has a net tax overpayment, the department shall provide the findings of this overpayment to the Comptroller for repayment of funds paid into the State Treasury through error pursuant to s. 215.26.

According to s. 212.12(6)(b), F.S., if a taxpayer does not have adequate records of sales or purchases, the Department may project a deficiency from a sample of the taxpayer's available records. There is no provision in s. 212.12(6)(b), F.S. to project an overpayment for a taxpayer who does not have adequate records.

Defining the Population

A stratified statistical sample can be performed on sales, purchases, or both. If certain events occurred during the audit period, such as a change in accounting systems or staff, a change in accrual methods for expensed purchases, new locations, and new product lines, then more than one population may be defined. There is no requirement for all transactions for the entire audit period to be contained in one population.

For a statistical sample performed on sales, the sales data population may include sales information for all customers, or only shipments made to Florida locations. When a "ship to" field is available, the data may be filtered to include only shipments to Florida locations. The Department attempts to eliminate shipments made to states other than Florida when defining the population.

For a statistical sample performed on expense purchases, the purchase data population may include all expense accounts or only selected expense accounts. The determination as to which to include may be made based on a review of the chart of accounts. The accounts payable file may include expense accounts for exempt services, taxes, insurance, etc. Accounts that do not have any tax implications may be removed from the population. A review of the general ledger accounts may be made to eliminate as many accounts as possible when defining the population.

The dollar amount that is the unit of measure is the dollar amount field contained in the electronic data. Sometimes the data will include line item amounts instead of invoice totals. In these cases, the line item amount would become the unit of measure.

Stratification

Stratified statistical sampling is the preferred method of sampling electronic records, as it divides the population into more homogeneous groups. Generally, seven strata is the optimal number necessary for any population. The Department uses seven sample strata for any statistically sampled population, unless the variability of the data indicates the population can be sampled without stratifying.

Population parameters are calculated by Department software. These parameters include measures of central tendency (mean, median, and mode) and measures of variability (range, variance, and standard deviation). The variability of the data indicates whether stratification is needed. Populations are generally stratified unless the variability is small.

A detail stratum is usually identified and separated from the population to be sampled. The detail stratum contains the high dollar items in the population. Generally speaking, such items represent a large portion of total dollars and only a small percentage of total items. The detail stratum is not sampled and not included in any projection made from the sample. A separate exhibit is prepared in the audit workpapers for the detail stratum. The detail stratum may consist of both negative and positive amounts from the data.

The taxpayer and the auditor negotiate where to set the dollar amount of the detail threshold. The decision of where to set the threshold and how many records will be examined in detail may be based on the estimated amount of time involved in performing the detail examination. The lower the threshold, the more records will be detailed and the fewer records will be sampled.

Sampling Agreement and
Stratified Statistical Sampling Plan

Section 212.12(6)(c), F.S., requires the Department to make a good faith effort to reach an agreement as to the means and methods to be used in sampling. Thus, a *Sampling Agreement* is presented to the taxpayer for signature (see Appendix 2). In the event no agreement is reached, the taxpayer is entitled to a review by the executive director, as provided in s. 212.12(6)(c), F.S. The *Stratified Statistical Sampling Plan* is attached to the *Sampling Agreement* (see Appendix 3). Each stratified statistical sample in the audit requires a separate stratified statistical sampling plan. The stratified statistical sampling plan provides the following information:

- Stratum boundaries, including population dollars and number of items per stratum.
- Method of selecting the pilot sample.
- Pilot sample size, including sample dollars and number of items per stratum.

Stratum boundaries are determined by Department software using the cumulative square root of the frequency method. This method marks off equal increments on the cumulative square root scale, calculates a point interval, and determines stratum boundaries using the point interval (see Appendix 4).

A pilot sample from each stratum is selected and analyzed to determine the variability of the data. The variability of the data is measured by the standard deviation of the additional tax due or overpaid from the pilot sample. The size of the pilot sample may vary for each statistical sample. Three alternative methods of determining pilot sample size are:

- Central limit theorem.
- Taxpayer request.
- Sample size formula.

The central limit theorem applies when a sample size of at least 30 sample points is used. In accordance with the central limit theorem, the sample means will be normally distributed for random samples taken from the population regardless of the population's distribution or size. Therefore, a pilot sample of 30 sample points per stratum will be used if this method is selected.

A sample size may be requested by the taxpayer. In these cases, the Department randomly selects the requested number as a pilot sample, allocating the number requested per stratum based on dollars or number of units. Regardless of which pilot sample method is chosen, a minimum sample size of 30 sample points per stratum is used by the Department, based on the central limit theorem.

Based on statistical sampling methodology, the formula used to calculate sample size is:

$$n = (ZS/E)^2$$

n = sample size
Z = normal deviate
S = sample standard deviation
E = acceptable magnitude for error

The following is an example using the formula n = $(ZS/E)^2$ to calculate an estimated sample size for stratum 1.

EXAMPLE

Sample size for stratum 1:

Z = 1.96
S = $34.026808
E = $3.75

n = $(ZS/E)^2$
n = $((1.96 * 34.026808) / 3.75)^2$
    $(17.784678)^2$
    317 records

The normal deviate (z) value at a 95% confidence level is 1.96. A table of z values is shown in Appendix 7. The standard deviation of the unit of measure (i.e., invoice or line item amount) in the electronic data is used in the formula, since the standard deviation of the audit exceptions (additional tax due or overpaid) is unknown until the pilot sample is performed. The standard deviation per stratum is calculated using Department software.

The acceptable magnitude for error (E) is the amount of sampling error agreed upon by the taxpayer and the Department, and it may vary from sample to sample. Every sample has a sampling error. A detail audit would have no sampling error, because all records in the population would be examined. E is a measure of how far the taxpayer and the Department are willing to deviate above or below the true, but unknown, mean error of the population.

In order to determine the pilot sample size, a value of E is calculated as a consistent percentage of each stratum midpoint. Department software defaults to a 6% starting point, which can be changed if needed.

<div style="border:1px solid">

EXAMPLE

Value of E for stratum 1:

Lower Boundary	Upper Boundary
$0.00	$125.00

Stratum midpoint = ($0.00 + $125.00) / 2
                              = $62.50
E = 6% * stratum midpoint
   = 6% x $62.50
   = $3.75

Therefore, the value of E for stratum 1 is $3.75 per sample point.

</div>

This step is repeated for each stratum. If the formula results in a number less than 30, the number is increased to 30 in accordance with the central limit theorem. All fractional results are rounded up, as partial records cannot be examined.

Pilot sample points are randomly selected from each stratum using Department software. In order to be statistically valid, every item in the population must have an equal chance of being selected in a random sample. Simple random sampling is used to select the sample points. A seed number, which will be large enough to avoid duplicate selection of sample points, is generated by Department software to begin the random selection. It is important to document the seed number used when selecting pilot sample points. If additional sample points are required for a final sample, the same seed number must be used to ensure that pilot sample points are not also selected as additional sample points. The seed number will be documented in the *Stratified Statistical Sampling Plan.*

Completing the Sample

The Department requests source documents to verify the tax treatment of the sample points. If tax was due on a transaction and not paid or accrued, an underpayment is scheduled in the audit workpapers. If tax was paid or accrued in error on a transaction and the taxpayer has met the requirements for a refund of taxes paid to the Department in error, an overpayment is scheduled. All sample points will be examined and documented in the audit workpapers. Correctly taxed transactions will not be deleted, but will hold a 0.00 additional tax due value.

Any sample points without supporting source documents are considered taxable and will be scheduled as an underpayment. The results of the pilot sample are provided to the taxpayer during the audit, so that any available supporting documentation can be located.

After the pilot sample has been completed, the final required sample size is calculated based on pilot sample results. Additional sample points, if needed, are randomly selected from the population. The formula used to calculate required sample size is:

$$n = (ZS/E)^2$$

> n = sample size
> Z = normal deviate
> S = sample standard deviation
> E = acceptable magnitude for error

Note that this is the same formula that may be used to calculate pilot sample size. The normal deviate (z) value at a 95% confidence level is 1.96. A table of z values is shown in Appendix 7. The standard deviation (S) of additional tax due or overpaid from the pilot sample is used to calculate the required sample size. A direct relationship exists between standard deviation and required sample size. As standard deviation increases, required sample size increases. Conversely, a smaller standard deviation will result in smaller required sample sizes.

The acceptable magnitude for error (E) is the amount of sampling error agreed upon by the taxpayer and the Department and may vary from sample to sample and audit to audit. E is a measure of how far the taxpayer and the Department are willing to deviate above or below the true, but unknown, mean error of the population. The value of E will be discussed with the taxpayer when determining the final sample size. Generally, the value of E is balanced with sample size to be cost effective yet acceptable. An inverse relationship exists between the value of E and required sample size. The larger the value of E, the smaller the sample size. The smaller the value of E, the larger the sample size.

Department software initially calculates a value of E based on the pilot sample results where no additional sample points are needed. Three alternative values of E are calculated at 25% increments that will require additional sample points.

Additional sample points may or may not be needed for each stratum, depending on the selected value of E. If additional sample points are required, they are randomly selected from the original population using Department software and the same seed number used to select the pilot sample. They are then added to the pilot sample transactions in the statistical exhibit and audited for tax compliance.

Projecting and Allocating Results

Sample results are projected per stratum. The results of the detail stratum are not included in the projection of additional tax due or overpaid from the sample, as the detail stratum is not sampled. Sample results are projected at the mean. Projecting at the mean is fair to both the taxpayer and the Department, because there is a 50% chance that the result is too high and a 50% chance that the result is too low.

Variable sampling provides an estimate of a population's characteristics expressed in dollars. Stratified statistical sampling is a form of variable sampling, because the result is a quantitative dollar estimate. With variable sampling, there is no minimum number of errors per stratum or in total. If one underpayment is found, it is projected as a deficiency. When a taxpayer's records are adequate for the audit period and requirements for a refund have been met, even one overpayment is projected as a credit. If no errors are found in a stratum, projected additional tax due is $0.00 for that stratum.

Three estimators may be used to project sample results:

- Difference estimator.
- Ratio estimator.
- Mean estimator.

The difference estimator requires a sample mean error (average error) to be calculated for each stratum (see Appendix 5). The sample mean error is the additional tax due or overpaid in each stratum divided by the sample size of each stratum. The sample mean error is multiplied by the total number of units in each stratum of the population and summed to obtain total additional tax due or overpaid for the population. Statistical literature identifies this estimator as the preferred estimator for projecting stratified statistical sample results.

The ratio estimator is an alternative to the difference estimator. The ratio estimator may be used if requested by the taxpayer. The ratio estimator requires that a percentage of error be calculated based on the ratio of additional tax due or overpaid from the sample divided by the dollars examined in each stratum. The ratio is then multiplied by population dollars in each stratum and summed to obtain total additional tax due or overpaid for the population.

The mean estimator requires recording tax due for each sample point. The sample mean tax due is the sum of the tax due for each sample point in the stratum divided by the number of sample points in each stratum. The sample mean tax due is multiplied by the total number of units in each stratum of the population and summed to obtain total tax due for the population. Tax remitted by the taxpayer for the period is then subtracted from this amount to determine additional tax due or overpaid.

The mean estimator can be used for taxpayers who elect to remit use tax using an alternative reporting method (remitting use tax based on a percentage of payables each month or a set amount of use tax each month). The mean estimator should be used in this situation, because tax paid on any one transaction cannot be verified.

Required sample size was calculated based on an acceptable magnitude for error (E) agreed upon by the taxpayer and the Department. Once the final sample has been audited and the results projected over the entire population, actual precision is calculated.

Precision is one-half of the range that may contain the projected amount at a given confidence level. The range is expressed as a plus or minus amount from the mean. The plus amount added to the mean is the upper confidence limit (UCL). The minus amount subtracted from the mean is the lower confidence limit (LCL).

The Department software uses the pooled approach when calculating precision on stratified statistical samples. This approach incorporates the variability of all strata and considers all strata in terms of their relative weights. Additional information, including precision formulas and examples of how the LCL and UCL are calculated for a sample using the pooled approach, is shown in Appendix 6. Department software computes precision at a 95% confidence level.

In order to calculate penalty and interest, projected sample results will be apportioned to monthly amounts. This may be based on the percentage of monthly purchases to total purchases, monthly sales to total sales, amounts reported on *Sales and Use Tax Returns* (Form DR-15), or other applicable records. The Department prefers to use the allocation method based on one of these sources when allocating results to monthly amounts. An alternative method to apportion sample results to monthly amounts is to use an average amount per month. This method should be used only when monthly sales or purchases data is not available. This method assumes the errors occurred uniformly throughout the audit period. Penalty and interest is calculated on additional tax due from the detail stratum based on the individual dates of the items examined in detail.

Appendix 1   Stratified Statistical Sampling Steps

Steps in Defining the Population

Step	Action	Considerations	Result
1	Determine the general audit area to be sampled.	• Sales. • Expense purchases.	The general audit area is identified and narrowed as much as possible.
2	Defining the most homogeneous population.	• Division(s). • Taxpayer location(s). • Accounting system changes. • Personnel changes.	The most homogeneous population is defined.
3	Remove adjusting journal entries.	• Adjusting journal entries. • Prepaid asset purchase transactions.	Adjusting journal entries are removed from the population; prepaid assets are added to the population or audited separately.
4	Identify the unit of measure.	• Total invoice amounts. • Line item amounts.	A dollar value is identified in the data to be used as the unit of measure.
5	Handle credit amounts.	• Match debits and credits, if practicable. • Absolute value remaining credits.	Matching debits and credits are removed from the population. All remaining credits are absolute valued.
6	Calculate population parameters.	• Population parameters including mean, median, standard deviation, skewness, and kurtosis are calculated with Department software.	The population to be sampled will be stratified into seven strata unless the range of the data is small.
7	Identify a detail threshold amount.	• Decide detail threshold amount. • Obtain agreement from the taxpayer.	Records to be examined in detail are identified.

Steps in Preparing the *Sampling Agreement* and

Stratified Statistical Sampling Plan (Attachment)

Step	Action	Considerations	Result
1	Determine stratum boundaries.	• Stratum boundaries optimally determined using the cumulative square root of the frequency method.	Generally, the population to be sampled is optimally stratified into 7 strata.
2	Determine pilot sample size.	• Central limit theorem. • Taxpayer requests a certain sample size (in total or per strata). • Formula.	A method to determine the pilot sample size is selected.
3	Select the pilot sample.	• Department software used to randomly select sample points using simple random sampling and a seed number. • Strata information is imported into the audit program as a statistical exhibit. • Sample points are imported into the audit program.	The pilot sample points to be examined are randomly selected using Department software. Strata information and sample points are imported into the audit program in a statistical exhibit.
4	Prepare *Sampling Agreement*.	• Standard Department *Sampling Agreement* is prepared.	*Sampling Agreement* is prepared to be presented to the taxpayer.
5	Prepare the *Sampling Agreement* attachment (*Stratified Statistical Sampling Plan*).	• *Stratified Statistical Sampling Plan* provides details of the sampling plan and is prepared by the auditor.	*Stratified Statistical Sampling Plan* is presented to the taxpayer along with the *Sampling Agreement;* every effort is used to obtain a signed *Sampling Agreement*.

Steps in Completing the Sample

Step	Action	Considerations	Result
1	Audit the pilot sample.	• Results are recorded in the audit exhibit. • Additional tax due or overpaid, as appropriate, is scheduled in the audit workpapers for the pilot sample points.	Audit results are documented and the standard deviation of exceptions per strata is obtained.
2	Calculate the required sample size at an agreed upon acceptable magnitude for error (E).	• Formula used is $n=(ZS/E)^2$ • Prepare *Required Sample Size Schedule*. • Identify a value to be used for E.	Required sample size is computed at various values of E, and taxpayer agreement is obtained on a specific value.
3	Select additional sample points required, if any.	• Department software is used to select sample points using simple random sampling. • Same seed number used to select the pilot sample is used to select additional sample points. • Additional sample points are imported into the statistical exhibit.	The additional sample points are randomly selected using Department software, scheduled in the audit workpapers, and examined by the auditor for tax compliance.

Steps in Projecting and Allocating Sample Results

Step	Action	Considerations	Result
1	Select an estimator to be used to project results.	• Difference estimator (preferred). • Mean estimator. • Ratio estimator.	The estimator to be used to project sample results is selected.
2	Project results using the selected estimator.	• Indicate selected estimator in audit program. • If mean estimator is selected, tax paid is entered and will be offset against tax due.	Additional tax due or overpaid is projected per strata using the selected estimator on the *Estimator Results Schedule*.
3	Compute precision.	• The LCL and UCL are computed by the audit program at a 95% confidence level.	LCL and UCL amounts are shown on the *Estimator Results Schedule*.
4	Allocate results to monthly totals.	• Compute a monthly percentage based on monthly amounts from taxpayer's data. • Compute a monthly percentage based on some other factor, like downloaded data from the *Sales and Use Tax Returns* (Form DR-15). • Compute an average tax due per month.	Additional tax due or overpaid is allocated to monthly amounts, for the purpose of computing penalty and interest.

Appendix 2   Sampling Agreement

This agreement, made and entered into by and between the Florida Department of

Revenue, hereafter referred to as "The Department", and <u>Company Name</u>,

FEI# _____, Audit #_____, hereafter referred to as "Taxpayer."

The Department and Taxpayer will be collectively referred to herein as "Parties."

Witnesseth

WHEREAS, the Parties wish to facilitate the completion of a sales and use tax

compliance audit, Audit Number _____;

WHEREAS, the Parties agree that the records of Taxpayer are so voluminous in nature

and substance that a sampling is an effective, expedient, and adequate method to facilitate the

audit;

NOW, THEREFORE, in consideration of the mutual premises contained herein, it is

hereby agreed to by the Parties as follows:

1. Pursuant to section 212.12(6)(c), Florida Statutes, the audit of Taxpayer for the

    period _____ shall be controlled by the sampling method set forth in

    attachment(s) hereto and specifically incorporated herein and made a part hereto.

2. This sample shall be deemed to be representative of all of the transactions pertaining

    to Taxpayer's business and the findings of this sampling shall be projected over the

    period set forth in the attachment.

In Witness whereof, we as representatives of the respective parties have hereunto set our hand and in so doing we hereby bind the parties who we represent.

Florida Department of Revenue

By:_____

    Jim Zingale, Executive Director or his designee(s)

_____

Title   Date

By:_____

    Officer or owner of business

_____

Title   Date

Appendix 3 (Example)   Stratified Statistical Sampling Plan (Sampling Agreement Attachment)

This attachment is made part of the sampling agreement, made and entered into by and between the Florida Department of Revenue, hereafter referred to as "The Department" and:

TAXPAYER NAME	AUDIT NO._____
	FEI #_____

        hereafter referred to as "Taxpayer."

1. Define the Population to be Sampled:

Florida Department of Revenue auditors will sample expense purchases. Adequate records are available for the period of 04/01/96 through 12/31/98, which will be the population to be statistically sampled. As provided under s. 212.12(6), F.S., any part of the audit period for which adequate records are not available will be sampled using a methodology other than stratified statistical sampling.

2. Stratification Plan:

Seven strata have been developed based on invoice dollar amounts. The method used for stratification is optimal and named the Cumulative Square Root of the Frequency Method. This method marks off equal intervals on the cumulative square root scale to determine strata boundaries. Credit values in the population are absolute valued for stratification purposes.

Strata #	Lower Boundary	Upper Boundary	Population # of Items	Population Dollars	Sample # of Items	Sample Dollars
1	0.00	125.00	2,341	165,291.99	30	2,161.34
2	125.03	250.00	1,464	269,709.30	30	5,583.22
3	250.25	474.73	1,005	347,103.69	30	10,268.84
4	475.40	750.00	747	447,962.17	30	18,426.70
5	751.00	1,124.00	589	539,915.06	30	27,735.27
6	1,128.00	1,724.86	406	577,272.99	30	42,705.52
7	1,725.14	2,983.75	235	509,694.49	30	65,989.55
		TOTAL	6,787	2,856,949.69	210	172,870.44

Exhibit B01 will include the sample points from the sample strata. A detail stratum consisting of records greater than $3,000 is agreed to be scheduled in a separate exhibit which will not be part of the stratified statistical sample exhibit.

3. Pilot Sample

A pilot sample of 210 sample points will be examined and the results used to determine the final sample size based on statistical theory and methodology. The pilot sample size of 210 was determined based on the central limit theorem, which applies when a sample size of at least 30 is used.

The central limit theorem states, regardless of the population's distribution or size, the sample means will be normally distributed for random samples taken from the population.

4. Random Selection of Sample Points

The pilot sample has been randomly selected from the population by the Department using Department software and the randomly generated seed number (see below). Taxpayer will provide the Department with the source documents for the sample points selected. The auditor determines additional tax due or overpaid for each pilot sample point.

Random Selection Seed Number: _____

5. Final Sample

After completion of the pilot sample, the variability of the data is determined and used to calculate the final sample size. The formula used to compute the final sample size is $n=(ZS/E)^2$ where n = the final sample size, Z = the normal deviate at a 95% confidence level (1.96), S = the standard deviation of the pilot sample, and E = sampling error.

Sampling error (E) is a measure of how far we are willing to deviate from the true, but unknown, average error of the population. The sampling error is balanced with the resulting sample size to be reasonable and acceptable. Sampling error is inversely proportional to sample size. As sampling error goes up, sample size goes down. The value to be used for E is discussed when calculating the final sample size.

After determining final sample size, if the required sample size is larger than the number examined in the pilot sample, additional sample points will be randomly selected by the Department using the same seed number as was used in the pilot. The taxpayer will provide these additional source documents to the Department for audit. The auditor determines

additional tax due or overpaid for each sample point. Any sample points for which the taxpayer cannot provide the source documents are considered taxable and an appropriate deficiency will be determined.

Because statistical sampling is based on the random selection of records from the entire population of records and because each record must have an equal probability of being selected, a statistical sample is representative of the entire population of records.

6. Calculation of Additional Tax Due or Overpaid

Additional tax due or overpaid will be calculated per stratum using the difference estimator. The Department assesses at the mean. Projecting at the mean is fair to both the Department and the Taxpayer, because there is a 50% chance that the result is too high and a 50% chance that the result is too low. There is no minimum number of errors. If no errors are found in a stratum, additional tax due or overpaid is 0.00 for that stratum. If even only a single error is found, it is projected, whether the error is a tax deficiency or overpayment.

The difference estimator calculates an average error per unit for each stratum (additional tax due or overpaid in the sample divided by the sample size). This average error is then multiplied by the total number of units in the population for each stratum and summed to obtain total additional tax due or overpaid for the population.

Total additional tax due or overpaid is allocated to each month within the exhibit period based on the percent of monthly totals to the total population amount for the exhibit. Penalty and interest, if applicable, will then be calculated on these monthly figures.

7. Sampling Plan Applicable Only to the Taxpayer

This *Sampling Plan* is only applicable to the Taxpayer that is a party to the *Sampling Agreement* for the particular audit or audit period noted herein, and is not intended to be a rule or a statement of general applicability.

Appendix 4 Cumulative Square Root of the Frequency Optimal Stratification Method

Cumulative Square Root of the Frequency Optimal Stratification Method

	Invoice Amts	Cumulative Frequency	Frequency	Square Root of Frequency	Cumulative Sq Rt of Frequency	
	0	0	0	0	0	
	25	8501	8501	92.20086767	92.20086767	
	50	14296	5795	76.12489737	168.325765	
	75	17358	3062	55.33534133	223.6611064	
	100	19148	1790	42.3083916	265.969498	
Strata 1	125	20637	1489	38.58756276	304.5570607	**20,637 Records**
	150	21844	1207	34.74190553	339.2989663	
	175	22943	1099	33.15116891	372.4501352	
	200	23781	838	28.94822965	401.3983648	
	225	24503	722	26.87005769	428.2684225	
	250	25124	621	24.91987159	453.1882941	
	275	25651	527	22.95648057	476.1447747	
	300	26212	561	23.68543856	499.8302132	
	325	26728	516	22.71563338	522.5458466	
	350	27139	411	20.27313493	542.8189815	
Strata 2	375	27493	354	18.81488772	561.6338693	**6,856 Records**
	400	27826	333	18.24828759	579.8821569	
	425	28170	344	18.54723699	598.4293938	
	450	28508	338	18.38477631	616.8141702	
	475	28789	281	16.76305461	633.5772248	
	500	29030	241	15.5241747	649.1013995	
	525	29265	235	15.32970972	664.4311092	
	550	29455	190	13.78404875	678.2151579	
	575	29654	199	14.10673598	692.3218939	
	600	29859	205	14.31782106	706.639715	
	625	30056	197	14.03566885	720.6753838	
	650	30256	200	14.14213562	734.8175194	
	675	30408	152	12.32882801	747.1463475	
	700	30547	139	11.78982612	758.9361736	
	725	30700	153	12.36931688	771.3054905	
	750	30874	174	13.19090596	784.4963964	
	775	31044	170	13.03840481	797.5348012	
	800	31183	139	11.78982612	809.3246273	
	825	31317	134	11.5758369	820.9004642	
	850	31513	196	14	834.9004642	
Strata 3	875	31658	145	12.04159458	846.9420588	**4,165 Records**
	900	31750	92	9.59166305	856.5337219	
	925	31860	110	10.48808848	867.0218104	
	950	31970	110	10.48808848	877.5098988	
	975	32059	89	9.43398113	886.94388	
	1000	32145	86	9.2736185	896.2174985	
	1025	32220	75	8.66025404	904.8777525	
	1050	32290	70	8.36660027	913.2443528	
	1075	32378	88	9.38083152	922.6251843	
	1100	32463	85	9.21954446	931.8447288	
	1125	32551	88	9.38083152	941.2255603	
	1150	32638	87	9.32737905	950.5529393	
	1175	32732	94	9.69535971	960.248299	
	1200	32849	117	10.81665383	971.0649529	
	1225	32918	69	8.30662386	979.3715767	
	1250	32982	64	8	987.3715767	
	1275	33060	78	8.83176087	996.2033376	
	1300	33138	78	8.83176087	1005.035098	
	1325	33188	50	7.07106781	1012.106166	
	1350	33254	66	8.1240384	1020.230205	
	1375	33316	62	7.87400787	1028.104213	
	1400	33382	66	8.1240384	1036.228251	

	1425	33432	50	7.07106781	1043.299319	
	1450	33503	71	8.42614977	1051.725469	
	1475	33547	44	6.63324958	1058.358718	
	1500	33615	68	8.24621125	1066.604929	
	1525	33682	67	8.18535277	1074.790282	
	1550	33731	49	7	1081.790282	
	1575	33789	58	7.61577311	1089.406055	
	1600	33843	54	7.34846923	1096.754524	
	1625	33896	53	7.28010989	1104.034634	
	1650	33944	48	6.92820323	1110.962838	
Strata 4	1675	33981	37	6.08276253	1117.0456	**2,323 Records**
	1700	34025	44	6.63324958	1123.67885	
	1725	34064	39	6.244998	1129.923848	
	1750	34112	48	6.92820323	1136.852051	
	1775	34152	40	6.32455532	1143.176606	
	1800	34201	49	7	1150.176606	
	1825	34252	51	7.14142843	1157.318035	
	1850	34278	26	5.09901951	1162.417054	
	1875	34317	39	6.244998	1168.662052	
	1900	34361	44	6.63324958	1175.295302	
	1925	34405	44	6.63324958	1181.928551	
	1950	34441	36	6	1187.928551	
	1975	34488	47	6.8556546	1194.784206	
	2000	34534	46	6.78232998	1201.566536	
	2025	34573	39	6.244998	1207.811534	
	2050	34601	28	5.29150262	1213.103037	
	2075	34635	34	5.83095189	1218.933988	
	2100	34683	48	6.92820323	1225.862192	
	2125	34715	32	5.65685425	1231.519046	
	2150	34774	59	7.68114575	1239.200192	
	2175	34811	37	6.08276253	1245.282954	
	2200	34849	38	6.164414	1251.447368	
	2225	34868	19	4.35889894	1255.806267	
	2250	34890	22	4.69041576	1260.496683	
	2275	34922	32	5.65685425	1266.153537	
	2300	34945	23	4.79583152	1270.949369	
	2325	34982	37	6.08276253	1277.032131	
	2350	35001	19	4.35889894	1281.39103	
	2375	35023	22	4.69041576	1286.081446	
	2400	35052	29	5.38516481	1291.466611	
	2425	35071	19	4.35889894	1295.82551	
	2450	35099	28	5.29150262	1301.117012	
	2475	35125	26	5.09901951	1306.216032	
	2500	35139	14	3.74165739	1309.957689	
	2525	35158	19	4.35889894	1314.316588	
	2550	35171	13	3.60555128	1317.922139	
	2575	35187	16	4	1321.922139	
	2600	35205	18	4.24264069	1326.16478	
	2625	35234	29	5.38516481	1331.549945	
	2650	35249	15	3.87298335	1335.422928	
	2675	35273	24	4.89897949	1340.321908	
	2700	35286	13	3.60555128	1343.927459	
	2725	35301	15	3.87298335	1347.800442	
	2750	35319	18	4.24264069	1352.043083	
	2775	35335	16	4	1356.043083	
	2800	35347	12	3.46410162	1359.507185	
	2825	35359	12	3.46410162	1362.971286	
	2850	35375	16	4	1366.971286	
	2875	35391	16	4	1370.971286	
	2900	35403	12	3.46410162	1374.435388	
	2925	35417	14	3.74165739	1378.177045	
	2950	35430	13	3.60555128	1381.782597	
	2975	35439	9	3	1384.782597	
	3000	35468	29	5.38516481	1390.167761	
Strata 5	3025	35476	8	2.82842712	1392.996189	**1,495 Records**
	3050	35489	13	3.60555128	1396.60174	
	3075	35495	6	2.44948974	1399.05123	

Continued at equal intervals
From 3,075 to 24,875…

	24875	36488	0	0	1947.280448
	24900	36488	0	0	1947.280448
	24925	36488	0	0	1947.280448
	24950	36488	0	0	1947.280448
	24975	36489	1	1	1948.280448
Strata 7	25000	36489	0	0	1948.280448	**376 Records**

36,489 Total Records

	278.3257783	**Point Interval**
	556.6515566	**Point Interval * 2**
Cumulative Sq Rt Frequency	1948.280448	834.9773349 **Point Interval * 3**
Divided by # of strata	7	1113.303113 **Point Interval * 4**
		1391.628892 **Point Interval * 5**
Equals the Point Interval	278.3257783	1669.95467 **Point Interval * 6**
		1948.280448 **Point Interval * 7**

In order to determine stratum boundaries, the point interval is calculated. Department software defaults at $25 increments but can be modified if needed. The point interval is 278.3257783, which is the last amount in the cumulative square root column (1948.280448) divided by the number of strata used. The point where the cumulative square root amount is at least 278.3257783 will be where the first stratum ends and the second stratum begins. The second stratum ends at the next point interval, which is 556.6515566 (278.3257783 x 2), and continues until all stratum boundaries are identified.

The number of units (or frequency) in each stratum is calculated below. The frequency of stratum 1 is 20,637 and is equal to the cumulative frequency. The frequency of stratum 2 is 6,856 units (the cumulative frequency of 20,637 at stratum 1 is subtracted from 27,493, the cumulative frequency at stratum 2). This process continues until all frequencies have been identified. The resulting stratum boundaries and frequencies for this example are:

Stratum	Lower Boundary	Upper Boundary	Number in Stratum
1	$ 0.00	$ 125.00	20,637
2	125.01	375.00	6,856
3	375.01	875.00	4,165
4	875.01	1,675.00	2,323
5	1,675.01	3,025.00	1,495
6	3,025.01	6,850.00	637
7	6,850.01	25,000.00	376
	Total number of records		36,489

Appendix 5 Difference Estimator Results Schedule

(Example)

Taxpayer: Sample Audit Company

FEI/SS #: 59-1234567

Exhibit Type: Statistical

Audit #A0012345678

Period: 01/01/1996 through 12/31/1999

Exhibit B01: Expenses - Difference Estimator

Taxes Under Chapter 212, F.S.

Strata	Lower Boundary	Upper Boundary	Population Items	Sample Items	Tax Errors	Avg. Error	Additional Tax Due
1	0.00	150.00	2,094	324	$824.19	2.543796	$5,326.71
2	150.40	600.00	603	129	2,785.64	21.594109	13,021.25
3	601.16	1,298.40	376	53	3,246.16	61.248302	23,029.36
4	1,302.00	2,240.00	288	30	3,345.27	111.509000	32,114.59
5	2,250.60	3,472.43	230	30	5,384.26	179.475333	41,279.33
6	3,478.00	5,920.12	155	30	8,764.54	292.151333	45,283.46
7	5,998.75	9,963.20	122	30	14,137.56	471.252000	57,492.74
Totals			3,868	626	$38,487.62		$217,547.44

LCL: $213,161.92

UCL: $221,932.96

Appendix 6   Calculation of Precision

<u>Legend</u>   *<u>Confidence Interval Computations</u>*

μ = Mean of the populations

N_i = Number of population elements in Stratum i, i = 1, 2, 3,...,L

N = Total number of elements in the population
       = $N_1 + N_2 + N_3 + ... + N_L$

n_i = Number of elements in the sample selected from stratum i, i = 1, 2, 3,...,L

x_i = Value of samplepoint

$\overline{x_i}$ = Mean of the sample selected from stratum i, i = 1, 2, 3,...,L

$\overline{X_i}$ = $\sum_{i=1}^{Ni} X_i / Ni$ = True mean -- The numerator would be the true total for the i-th stratum

$\overline{x_i}$ = $\sum_{i=1}^{L} x_i / ni$ = Sample mean for the i-th stratum

W_i = N_i / N = Weight for the i-th stratum

σ^2 = Variance of the population measurements

σ = Population standard deviation

s^2_i = Variance of the sample measurements from stratum i, i = 1, 2, 3,...,L

s = standard deviation based on sample results

$\int$ = n_i / N_i = Proportion of invoices in the sample from the i-th stratum -- AKA "sampling fraction"

p = Population proportion

$\hat{p}_i$ = stratum estimate of population proportion

$\hat{q}$ = $(1-p)$ $\hat{q} = (1-\hat{p})$

τ = Population total

$\hat{\tau}$ = Estimate of τ (population total) based on sample results

An Approximate 95% Confidence Interval for the Population Mean μ :
Stratified Random Sampling (Difference Estimator)

$$\overline{X}_{st} \pm 1.96 \sqrt{\frac{1}{N^2} \left\{ \sum_{i=1}^{L} N_i \left[(N_i - n_i) / N_i \right] s^2_i / n_i \right\}}$$

where,

$$\overline{X}_{st} = 1/N \left(N_1 \overline{x}_1 + N_2 \overline{x}_2 + ... + N_L \overline{x}_L \right) = 1/N \sum_{i=1}^{L} N_i \overline{x}_i$$

This is our estimate of the population mean μ based on the stratified random sample. When the sample size n_i is small relative to the stratum size ($n_i < 5\%$ of N_i) this formula becomes:

$$\overline{X}_{st} \pm 1.96 \sqrt{\frac{1}{N^2} \left\{ \sum_{i=1}^{L} N_i s^2_i / n_i \right\}}$$

Numerical example:

i Stratum	N_i Population Size/Stratum	n_i Sample Size	$\overline{x}_i$ Sample Mean	s^2_i Stratum Sample Variance
1	13,513	128	3.08	2.04
2	33,421	341	4.12	2.38
3	32,781	248	3.93	3.42
4	18,941	161	4.04	5.98

$$N = \sum_{i=1}^{L} N_i = 98,656$$

Stratified mean =

$$\overline{X}_{st} = 1/N \left(N_1 \overline{x}_1 + N_2 \overline{x}_2 + N_3 \overline{x}_3 + N_4 \overline{x}_4 \right) =$$
$$1/98,656 \left\{ (13,513)(3.08)+(33,421)(4.12)+(32,781)(3.93)+ (18,941)(4.04) \right\}$$

$$= 3.899$$

Confidence interval for μ is:

$$\overline{X}_{st} \pm 1.96 \sqrt{ 1/N^2 \left\{ \sum_{i=1}^{L} N_i^2 \, s_i^2 / n_i \right\} }$$

$$3.90 \pm 1.96 \sqrt{ 1/98,656^2 \left\{ 13,513^2(2.04 / 128) + 33,421^2(2.38 / 341) + 32,781^2(3.42 / 248) + 18,941^2 (5.98 / 161) \right\} }$$

$$3.90 \pm 1.96 \left[\sqrt{.002623} = .051211 \right] = 3.90 \pm .100373$$

THUS LCL = 3.798686 $\Rightarrow$ UCL = 4.00373

An Approximate 95% Confidence Interval for the Population Total τ: Stratified Random Sampling (Mean Estimator)

$$\hat{\tau} \pm 1.96 \sqrt{\left\{ \sum_{i=1}^{L} N_i^2 \left[(N_i - n_i) / N_i \right] s^2_i / n_i \right\}}$$

where,

$$\hat{\tau} = N \, \overline{x}_{st}, \text{ and}$$

$$\overline{X}_{st} = 1/N \left(N_1 \overline{x}_1 + N_2 \overline{x}_2 + ... + N_L \overline{x}_L \right) = 1/N \sum_{i=1}^{L} N_i \overline{x}_i$$

When the sample size n_i is small relative to the stratum size ($n_i < 5\%$ of N_i) this formula becomes:

$$\hat{\tau} \pm 1.96 \sqrt{\left\{ \sum_{i=1}^{L} N_i^2 \, s^2_i / n_i \right\}}$$

Numerical example:

i Stratum	N_i Population Size/Stratum	n_i Sample Size	$\overline{x}_i$ Sample Mean	s^2_i Stratum Sample Variance
1	4,149	408	149	48,764
2	2,872	210	186	50,328
3	3,107	240	190	48,640

$$N = \sum_{i=1}^{L} N_i = 10,128$$

The confidence interval for $\hat{\tau}$ is

$$\hat{\tau} \pm 1.96 \sqrt{\left\{ \sum_{i=1}^{L} N_i^2 \left[(N_i - n_i) / N_i \right] s_i^2 / n_i \right\}}$$

or

$$\hat{\tau} \pm 1.96 \sqrt{\left\{ \sum_{i=1}^{L} N_i^2 s_i^2 / n_i \right\}}$$

if $n_i < 5\%$ of N_i

where,

$$\hat{\tau} = N \bar{x}_{st} = N \left\{ 1/N(N_1 \bar{x}_1 + N_2 \bar{x}_2) \right\}$$

$$= N_1 \bar{x}_1 + N_2 \bar{x}_2 + N_3 \bar{x}_3 = (4{,}149)(408) + (2{,}872)(210) + (3{,}107)(240) = 3{,}041{,}592$$

Substituting into the formula for the confidence interval yields:

$$(3{,}041{,}592) \pm 1.96 \sqrt{ \begin{aligned} &(4{,}149)^2 \left\{ ((4{,}149 - 408) / 4{,}149)(48{,}764 / 408) \right\} + \\ &(2{,}872)^2 \left\{ ((2{,}872 - 210) / 2{,}872)(50{,}328 / 210) \right\} + \\ &(3{,}107)^2 \left\{ ((3{,}107 - 240) / 3{,}107)(48{,}640 / 240) \right\} \end{aligned} }$$

or, $3{,}041{,}592 \pm 145{,}260.50$

LCL=3,041,592–145,260.50=**2,896,331.50;** *UCL*=3,041,592+145,260.50=**3,186,852.50**

Approximate 95% Confidence Interval for a Population Proportion p : Stratified Random Sampling (Ratio Estimator)

$$\hat{p}_{st} \pm 1.96 \sqrt{1/N^2 \left\{ \sum_{i=1}^{L} N_i^2 \left[(N_i - n_i)/Ni \right] \hat{p}_i \hat{q}_i / n_i - 1 \right\}}$$

where,

$$\hat{p}_{st} = 1/N (N_1\hat{p}_1 + N_2\hat{p}_2 + ... + N_L\hat{p}_L) = 1/N \sum_{i=1}^{L} N_i \hat{p}_i$$

is the estimate of p based on the stratified random sample and $\hat{p}_i$ is the sample proportion from stratum i, i = 1, 2, 3,...,L

When the sample size n_i is small relative to the stratum size ($n_i < 5\%$ of N_i) this formula becomes:

$$p_{st} \pm 1.96 \sqrt{1/N^2 \left\{ \sum_{i=1}^{L} N_i^2 \hat{p}_i \hat{q}_i / n_i - 1 \right\}}$$

A numerical example:

i Stratum	N_i Number in stratum	n_i Sample Size	$\hat{p}_i$ Stratum Sample Proportion (% of Error)
1	13,513	128	.26
2	33,421	341	.19
3	32,781	248	.30
4	18,941	161	.34

N = 98,656

$$\hat{p}_{st} = 1/98,656 \left\{ (13,513)(.26) + (33,421)(.19) + (32,781)(.30) + (18,941)(.34) \right\}$$
$$= .264937$$

Using this formula, because of small sampling ratios

$$\hat{p}_{st} \pm 1.96 \sqrt{1/N^2 \left\{ \sum_{i=1}^{L} N_i^2 \, \hat{p}_i \, \hat{q}_i / n_i - 1 \right\}}$$

we get:

$$.264937 \pm 1.96 \sqrt{ \begin{aligned} 1/98,656^2 \Big\{ (13,513)^2 \, [(.26)(.74)/127] \; + \\ (33,421)^2 \, [\,(.19)(.81)/340] \; + \\ (32,781)^2 \, [\,(.30)(.70)/247] \; + \\ (18,941)^2 \, [\,(.34)(.66)/160] \Big\} \end{aligned} }$$

or, $.264937 \pm .02946$

Or,   **LCL = .235476 ➜ UCL = .294398**

Appendix 7   Table of Z Values

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4014
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4725	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987									
3.5	.4997									
4.0	.4999									

Example: for z = 1.96, shaded area is 0.4750 out of the total area of 1.0000

Reprinted from Schaum's Outlines, <u>Business Statistics</u>, Third Edition.

GLOSSARY

Acceptable magnitude for error (E)
The level of acceptable risk associated with using the sample mean to estimate the population mean. Also known as precision, maximum tolerable error, sampling error, and E.

Average
See mean.

Central limit theorem
States that the sampling distribution of sample means approaches the normal distribution as the number of samples increases, regardless of the distribution of the population. If the central limit theorem is to be used, the samples drawn from the parent population must be of size 30 or more.

Confidence interval
The range of values in which the value of the population parameter of interest may be contained with the specified probability.

Confidence level
The specified probability of the confidence interval.

E
See acceptable magnitude for error.

Frequency distribution
A table listing possible values for a variable by class and their frequency.

Kurtosis
A descriptive characteristic of the values' distribution in a frequency curve. Kurtosis gives a measure of the peakedness of the distribution.

Machine-Sensible Data
A collection of related information in an electronic format. Machine-sensible records do not include hard copy records that are created or recorded on paper or stored in (or by) an imaging system such as microfilm, microfiche, or storage-only imaging systems.
Maximum tolerable error

See acceptable magnitude for error (E).

Measures of central tendency
Values calculated for a group of data and used to identify location or amounts. Measures include the mean, median and mode.

Measures of dispersion
Values calculated for a group of data and used to describe the variability of the data. Measures include variance, standard deviation, skewness and kurtosis, to name a few.

Measures of location
See measures of central tendency.

Measures of variability
See measures of dispersion.

Mean
A measure of central tendency calculating the sum of the values in the data group divided by the number of values.

Median
A measure of central tendency equal to the value of the middle item in a group of values when the values in the group are arranged in ascending or descending order.

Mode
A measure of central tendency equal to the value that occurs most frequently in a group of values.

Normal deviate
The normal deviate is represented by z. A table of Z values (Appendix 7) presents the probability that the value of interest lies within a specified interval under the unit normal distribution as measured by the standard deviation.

Parameter
A descriptive measurement of the population which is compiled based on each and all items which are members of the population of interest.

Point estimates
Numeric values of a sample statistic used to estimate the value of a population parameter.

Population
The entire set of values that is being audited.
Population parameters

The measured characteristics of a population.

Population variance
A measure of dispersion calculating the squared difference between each value in the population and the mean of the population divided by the population size. The square root of the population variance is the population standard deviation.

Precision
See acceptable magnitude for error (E).

Random number sampling
A table of numbers used to specifically select population elements for sample purposes.

Random sample
A type of sampling in which every value in the population has an equal chance of selection for inclusion in the sample.

Range
The difference between the highest and lowest values in a group of values.

Sample
A group of values selected from a population.

Sample variance
A measure of dispersion calculating the squared difference between each value in the sample and the sample mean divided by the sample size.

Sample point
A single record in the population selected to be audited. The record is the unit of measure in the taxpayer's data and can be an invoice amount, invoice line item amount, or other amount. Also referred to as sample unit.

Sample precision
The accuracy of the sample to represent the population.

Sampling error
See acceptable magnitude for error (E).

Sample unit
See sample point.

Skewness
A descriptive characteristic of the values' distribution in a frequency curve. A distribution can be negatively skewed (if the mean is greater than the median) or positively skewed (if the mean is less than the median) or symmetrical (if the mean is equal to the median).

Standard deviation
A measure of dispersion that is calculated by taking the square root of the variance.

Statistic
The estimate of a parameter derived from an unbiased and best estimator based on a random sample.

Statistical sampling
Using approaches to make inferences about a population on the basis of information obtained from a random sample.

Strata
Multiple subgroups of values with similar characteristics.

Stratum
A single subgroup of values with similar characteristics.

Stratification
The process of dividing the population into strata.

Variance
See population variance and sample variance.

Unit normal distribution
A continuous probability distribution that is symmetrical. The mean, median and mode are all equal.

Universe
See population.

Z
See normal deviate (see Appendix 7).

Bibliography

Arkin, Herbert. (1974). *Handbook of Sampling for Auditing and Accounting,* McGraw-Hill Book Company, New York.

Armitage, P. (1947). A comparison of stratified with unrestricted random sampling from a finite population. *Biometrika,* **34,** 273-280.

Chatterjee, S. (1967). A note on optimum stratification. *Skand. Akt.,* **50,** 40-44.

Chatterjee, S. (1972). A study of optimum allocation in multivariate stratified surveys. *Skand. Akt.,* **55,** 73-80.

Cochran, W. G. (1961). Comparison of methods for determining stratum boundaries. *Bull. Int. Stat. Inst.,* **38,** 2, 345-358.

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley and Sons , New York, third edition.

Cornfield, J. (1951). The determination of sample size. *Amer. Jour. Pub. Health,* **41,** 654-661.

Dalenius, T., and Gurney, M. (1951). The problem of optimal stratification II. *Skand. Akt.,* **34,** 133-148.

Dalenius, T., and Hodges, J. L., Jr. (1959). Minimum variance stratification. *Jour. Amer. Stat. Assoc.,***54,** 88-101.

David, F. N., and Neyman, J. (1938). Extension of the Markoff theorem of least squares. *Stat. Res. Mem.,* **2,** 105.

David, I. P., and Sukhatme, B. V. (1974). On the bias and mean square error of the ratio estimator. *Jour. Amer. Stat. Assoc.,* **69,** 464-466.

Erdos, P., and Renyi, A. (1959). On the central limit theorem for samples from a finite population. *Pub. Math. Inst. Hungarian Acad. Sci.,* **4,** 49-57.

Evans, W. D. (1951). On stratification and optimum allocations. *Jour. Amer. Stat. Assoc.,* **46,** 95-104.

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications,* John Wiley and Sons, New York, second edition.

Guy, Dan M.; Carmichael, Douglas R.; and Whittington, O. Ray. *Audit Sampling-An Introduction.* John Wiley and Sons, New York, Fourth Edition.

Lapin, Lawrence L. (1973). *Statistics For Modern Business Decisions,* Harcourt Brace Jovanovich, Inc. New York.

Mendenhall, William, Beaver, Robert J., and Beaver, Barbara M. *A Course in Business Statistics,* Duxbury Press, Belmont, California, Fourth Edition.

Neter, J. (1972). How accountants save money by sampling. *Statistics, A Guide to the unknown,* J. M. Tanur et al. (Eds), Holden-Day, Inmc., San Francisco, 203-211.

Wilburn, Arthur J.(1984). *Practical Statistical Sampling For Auditors,* Marcel-Dekker, Inc. Zikmund, William G. (1991). *Business Research Methods.* Harcourt Brace Jovanovich, Inc. New York.

FLORIDA

DEPARTMENT
OF REVENUE

GT-300034
R. 05/02