

Selecting a representative sample and the importance of narrowing the sampling frame to taxable items of interest

Dr. Al Kvanli

This material is the property of the presenter and cannot be reproduced or used without the expressed written consent of the presenter.

III-1

A clear definition of what a representative sample means statistically

- The easiest way to obtain a representative sample is to obtain a simple or stratified random sample from the sample frame using a "certified" random number generator.
- Since no portion of the sample frame is ignored or given special treatment, it provides a "fair" representation of the frame.
- The generally accepted method of certifying a random number algorithm is to subject it to a battery of tests called DIEHARD, developed by G. Marsaglia at Florida State University.

III-2

A clear definition of what a representative sample means statistically

- Other sample selection methods exist, whereby the frame is first divided into subgroups (e.g., in-state and out-of-state). Random samples can then be obtained from each of the subgroups.
- Since this amounts to stratified sampling, care must be taken that the sample size for each subgroup (stratum) satisfies the state's minimum sample size per stratum policy.

III-3

A clear definition of what a representative sample means statistically

- If the frame is to be divided into subgroups, it is important that this be carried out by the taxpayer and auditor before that actual audit takes place.
- This will require a careful examination of the sample frame by the taxpayer and auditor, followed by a detailed discussion as to exactly how the sample items will be selected.

III-4

Distinguishing between bias and reliability

- The *reliability* of an estimated population total refers to how reproducible the estimator is over repetitions of the sampling procedure.
- Assuming no measurement error in the sample, the reliability of an estimator can be measured by its *standard error*.
- The smaller the standard error, the greater is its reliability.

III-5

Bias of an Estimator

- The *bias* of an estimated population total refers to how the mean of the estimator over repetitions of the sampling procedure differs from the true population total.
- An **unbiased** estimator of a population total is one where, if sampling indefinitely, the average of the estimates will equal the true population total.

III-6

Accuracy of an Estimator

- The *accuracy* of an estimator of a population total refers to how far away a particular value of the estimate is, on average, from the true population total.
- The accuracy of an estimator can be determined by computing its *mean square error* (MSE), where
$$\text{MSE} = (\text{standard error})^2 + (\text{bias})^2$$

III-7

Accuracy of an Estimator

- The smaller the MSE of an estimator, the greater is its accuracy.
- Ratio and regression estimators very often have good accuracy even though they are slightly biased. In such situations, the estimated standard error is smaller (often much smaller) than that obtained using the difference estimator.

III-8

The Bathroom Scale Nobody Wants

- Consider a bathroom scale that consistently overweighs you by 10 lbs. each time it is used.
- The results obtained from this scale are
 - reliable (consistently overweighs you by 10 lbs.)
 - biased (your true weight is not what the scale indicates – the bias is 10 lbs.)
 - not accurate (due to the large bias)

III-9

Poor Precision v. Poor Validity

- On occasion, statistical estimates are labeled as “invalid” because the sample size was “too small” or the sampling frame was not stratified.
- Such comments are referring to the audit precision (reliability) rather than the validity of the result.
- Poor precision has nothing to do with the validity of the sample projection.

III-10

The dangers of using sub-sample proportions or means to determine if it is a representative sample

- Situation: When examining a simple random sample of size 100, it was pointed out to the auditor that only 2 of the sample transactions (2%) were marked as tax exempt.
- The sample frame consists of 5000 items.
- The taxpayer claims that in the sample frame, 6% of the transactions (300 transactions) were tax exempt and therefore the sample is not representative.

III-11

The dangers of using sub-sample proportions or means to determine if it is a representative sample

- Discussion: Let p be the proportion of frame items that are tax exempt.
- Based on the sample results, the estimate of p is $2/100 = .02$ (2%).
- An exact 90% confidence interval for p (based on the hypergeometric distribution) is from .38% to 6.12%.
- Consequently, based on the sample results, we are 90% confident that p is between these two values.
- But this interval includes 6% and so the sample results are consistent with the claim that p is 6%.

III-12

The dangers of using sub-sample proportions or means to determine if it is a representative

- Conclusion: Very often, sample proportions (or means) differ from the corresponding population proportions (or means) because of sampling error and not because the sample lacks representativeness.

III-13

Properly scrutinizing the chart of accounts and the electronic transactions in order to narrow the target population to items of interest

- In defining *items of interest*, ultimately the auditor is attempting to reduce the number of items in the population that contain the value of zero in the examined value.
- Items of interest can be, depending on the audit/sample objective, those transactions upon which tax is:
 - a) due regardless of whether it was paid, or
 - b) due and not paid, and/or
 - c) not due and was not paid

III-14

Properly scrutinizing the chart of accounts and the electronic transactions in order to narrow the target population to items of interest

- In a), the auditor wants to estimate the total subject to tax as he/she knows how much tax was paid.
- The auditor may wish to find the deficiencies only – in which case b) applies.
- A taxpayer may wish to make claims on tax overpayments, in which case c) would apply.
- Finally, an auditor might want to find the net overall tax adjustment, which means both b) and c) apply (the most typical use of statistical sampling).

III-15

Properly scrutinizing the chart of accounts and the electronic transactions in order to narrow the target population to items of interest

- Note in most all cases, regardless of the objective a), b), and/or c), we are not interested in items where no tax is due and no tax was paid.
- Audit experience can be used to scrutinize the accounts that likely contain items meeting the objective, or excluding accounts that likely contain only items *not of interest*
- In developing accounts that include items of interest, it often is an exercise of excluding accounts that contain items not of interest.

III-16

Properly scrutinizing the chart of accounts and the electronic transactions in order to narrow the target population to items of interest

- Sometimes downloads contain records that can be excluded:
 - Divisions and locations not audited,
 - Time periods outside the audit period,
 - Vendor information can be used to exclude certain accounts,
 - Dependent on jurisdiction statutes (what is taxable in one state may not be true in another),
 - Other items that with some initial investigation, shows that these items are not of interest.

III-17

Properly scrutinizing the chart of accounts and the electronic transactions in order to narrow the target population to items of interest

- When the auditor uses judgment to exclude certain accounts, and becomes overaggressive, the danger is that the auditor is not examining the right transactions.
- On the other hand, if the auditor is too conservative, and leaves in too many items that should have been excluded from the target population, precision suffers.

III-18

Sample Precision

- Sample precision can be measured by computing the
 - precision amount, or the
 - relative precision
- The precision amount is the amount added to and subtracted from the point estimate when deriving a confidence interval.
- The relative precision is the precision amount divided by the point estimate (expressed as a percentage).

III-19

Basic Rule

- As the error rate increases, the precision amount increases.
- As the error rate increases, the relative precision decreases.

III-20

More on Precision

- In the discussion to follow, we will focus on the relative precision.
- The relative precision is only meaningful if the population values have the same sign (all positive or all negative).
- If the population (and sample) contain positives and negatives, the point estimate may be near zero, whereby the relative precision is terrible yet the precision amount can be very good.

III-21

The importance of developing a sampling frame that constitutes taxable transactions to improve precision

- Every zero encountered in the sample will tend to increase the relative precision provided a few nonzero errors have been found.
- The lower the error rate (the more zeros found as a percentage of the sample), the larger the relative precision will be in the results.
- Restricting the sampling frame to the items of interest, and limiting the number of zeros that enter into the sample, is the best approach in assuring better a better relative precision.

III-22

Taxpayer Concerns

- One of the concerns of taxpayers is that auditors try to limit the sampling frame only to transactions where if errors exist, they are tax **underpayments**.
- Some taxpayers want the ability to define areas of interest where if an error exists, it is likely to be a tax **overpayment**.
- Many of the states are vague in their official policies.

III-23

Taxpayer Concerns

- Some of the field auditors and some of the computer audit specialists are more aggressive than others in trying to avoid tax overpayments in the sampling frame.
- The most common techniques for avoiding tax overpayments are avoiding transactions with sales tax paid to vendor, use tax accrued to the state, and certain areas where taxpayers often fail to take exemptions such as environmental and safety equipment.

III-24