STANDARD ON RATIO STUDIES

Exposure Draft May 2026

International Association of Assessing Officers

IAAO assessment standards represent a consensus in the assessing profession and have been adopted by the Board of Directors of the International Association of Assessing Officers (IAAO). The objective of IAAO standards is to provide a systematic means for assessing officers to improve and standardize the operation of their offices. IAAO standards are advisory in nature and the use of, or compliance with such standards is voluntary. If any portion of these standards is found to be in conflict with national, state, or provincial laws, such laws shall govern. Ethical and/or professional requirements within the jurisdiction may also take precedence over technical standards.

About IAAO

The International Association of Assessing Officers, formerly the National Association of Assessing Officers, was founded for the purpose of establishing standards for assessment personnel. IAAO is a professional membership organization of government assessment officials and others interested in the administration of the property tax. Over the years, IAAO members have developed assessment practice and administration standards. Many of these standards have been adopted by state and international oversight agencies, and some have been incorporated into legislation. IAAO continues at the forefront of assessment in North America and has been expanding its reach to the global community for the last five decades. Because standards form the rules by which North American assessors perform their duties, they may not be directly applicable to an overseas audience. The standards have been updated to also present the broad principles upon which the rules are based. IAAO believes those principles may be adapted to many differing statutory and regulatory scenarios worldwide.

Acknowledgements

This exposure draft is the work of the 2023-2026 Research and Standards Committees. While there were 25 individuals who participated during that time, we wish to offer special acknowledgement to the members who served for two or more years: Melissa Baer, CAE, Committee Chair, Alan Dornfest, AAS, FIAAO, Committee Vice Chair, Russ Thimgan, FIAAO, Michael Safarty, RES, Joshua Myers, FIAAO, Luc Hermans, Patrick Santoso, Marco Kuijper, Carmela Quintos, PhD, and Carol Neihardt. The final revisions are the work of the 2026 Committee, which also includes members Bill MacGougan, Brian Guerin, Dan Cypert, RES, Inhyuk Kwon and Michael Mills, CAE, RES, MAS, AAS.

This update would not have been possible without the expertise provided by Subject Matter Expert and honorary Committee Member, Robert Gloudemans. The Committee would also like to acknowledge the contributions of Garin Evans, Alex Chizewsky and Leah Thimgan in their support of the Committee's efforts.

At the time the project to revise the 2013 Standard on Ratio Studies was begun, the committee chair was Margie Cusack, FIAAO. Work continued under Jennifer Rearich, CAE, MAS chair during 2024 and part of 2025 and the draft was completed under Melissa Baer, CAE, chair from 2025 through the present. Additionally, we thank IAAO staff members Shannon Hiss, RES and Elizabeth Ferguson, PhD, Committee Liaisons, as well as Doug Warr, AAS, Board liaison, for their tremendous assistance along the way.

Revision Notes

This standard replaces the 2013 Standard on Ratio Studies and is a complete revision. The 2013 standard updated the 2010 standard.

Published by
International Association of Assessing Officers
314 W 10th St
Kansas City, Missouri 64105-1616
816-701-8100 • Fax: 816-701-8149
www.iaao.org
Library of Congress Catalog Card Number: ISBN

TABLE OF CONTENTS

1. SCOPE

This standard provides recommendations and guidance on the design, preparation, interpretation, and use of ratio studies for valuation purposes. It is divided into thirteen sections and includes a glossary of key terms. Many topics addressed in this standard are discussed in more detail in Property Appraisal and Assessment Administration (International Association of Assessing Officers (IAAO), 1990) and Mass Appraisal of Real Property (Gloudemans R. , 1999).

The primary users of this standard are those creating, using, or interpreting ratio studies for mass valuations intended for property tax assessment purposes. This includes professionals involved in the development or evaluation of valuations from models or other methods, as well as those in auditor or oversight roles evaluating the performance of an entire assessment jurisdiction. Additionally, the concepts within this standard are relevant for other mass valuation projects such as mortgage portfolio management.

IAAO recognizes there are multiple ways of referring to appraisers and appraisals throughout the world. Therefore, within this document the terms *valuer*, *assessing officer*, *lister*, and *appraiser* may be used interchangeably. In addition, the terms *valuation* and *appraisal* as used in this document are interchangeable and synonymous.

2. OVERALL PRINCIPLES

The key principles within this standard are as follows:

1. Ratio Studies are a fundamental part of the mass valuation process and should be used by both Assessors and Oversight Agencies to ensure valuations are fair and equitable.
2. Sample size and representativeness can have a significant impact on the interpretation of Ratio Study results.
3. Ratio studies may be done to establish the effectiveness and results from either a pure market value standpoint or a statutorily determined valuation system.
4. Interpretation may differ when using market values vs statutorily defined values (i.e. just value, use value, capped value, taxable value, gross assessment, or net assessment) and assumptions must be clear in order to interpret results. Use of any estimate other than market value changes the context of the statistical analysis.
5. Standards are intended to represent minimum acceptable performance. Assessors should strive for continuous improvement towards target levels and uniformity.

3. INTRODUCTION

Ratio studies are a statistical exercise integral to determining the valuation performance of an assessment jurisdiction's estimate of statutorily defined market value. A ratio consists of that estimate divided by a market value proxy, such as a sale price or independent property valuation. By extension, a ratio study consists of multiple ratios being evaluated to determine overall compliance with the various metrics described in this standard and is a critical tool for judging the validity of market-based valuations when compared to market-based proxies.

The use of non-market valuation estimates can mask the true level of valuation and confuse the measurement of valuation uniformity when the legal assessment level is other than 100 percent for all property or when statutory valuation constraints exist. Care should be taken to ensure a ratio study is properly specified and clearly identifies whether values are market or non-market based so that readers are clear about the purpose of the study, have reasonable expected outcomes, and understand statistically valid interpretations that can be made.

Principles

- Value must be defined and could be based on market or specific legal requirements.
- Market value cannot be observed directly; a single sales transaction does not equate to market value and ratio study statistics do not apply to individual properties.
- Ratio studies serve many purposes, and the purpose of the study will dictate how the statistics are analyzed and interpreted.

3.1. THE CONCEPT OF MARKET VALUE

Market value is the goal of most mass valuation assignments. The objective is to estimate the market value of properties based on legal requirements or accepted valuation definitions. When estimating market value, a group of sales will typically be utilized in the determination of the most probable price a willing buyer and willing seller would agree upon. A single market sale transaction, by itself, is not market value.

Market value is a concept in economic theory and cannot be observed directly. While a single sale may provide an indication of the market value of the property in question, a ratio study must be based on an adequate sample. Ratios are formed by dividing the appraised value by the sale price. The ratio can be multiplied by 100 and expressed as a percentage. Market values can be represented in ratio studies by valid arm's length sales prices. It is desirable that these sales have been reviewed, screened, and adjusted as necessary (Standard on verification and adjustment of sales, 2020). Sales prices provide the most objective estimates of market values and under normal circumstances should provide good indicators of market value. Therefore, the main concept of ratio studies is to systematically compare assessed values with sales prices as a starting point for calculating statistical measures to do in-depth analysis. Ratio studies may compare assessed values to market value proxies using independent appraisals. (See Appendix A- Independent Value Estimate-based Ratio Studies)

3.2. VALUATION ACCURACY

When mass valuation is used for property tax purposes, the credibility of the property tax depends largely on the accuracy of such value estimates to reflect market value as defined by professional standards and legal requirements. Therefore, the accuracy of valuations is of concern not only to assessors, but also taxing authorities, property taxpayers, and elected representatives.

There are two major components of accuracy: level and uniformity. Valuation level refers to the overall ratio of valuations to sales prices or market proxies. Level measurements provide information about the degree to which goals or certain legal requirements are met. Uniformity refers to the degree to which properties are appraised at equal percentages of those same sales prices or market proxies.

3.3. USES OF RATIO STUDIES

Ratio studies can be used for many purposes, both in the private and public sector by valuers, appeal boards, taxpayers, taxing authorities, oversight agencies, and others. Key uses of ratio studies are as follows:

- Measure and evaluate the level and uniformity
- Quality assurance and identify appraisal priorities
- Determine whether administrative or statutory standards have been met
- Determine time trends (See (Standard on verification and adjustment of sales, 2020))
- Test, refine, complete, and finalize values throughout the valuation process
- Direct and Indirect Equalization

It is important to note that ratio study statistics cannot be used to judge the level of valuation of an individual property. Such statistics can, however, be used to adjust assessed values on appealed properties to the common level.

3.3.1 Use by Oversight Agencies

National/State/provincial oversight agencies often perform ratio studies to advise assessing officers and the public about local market conditions and to order revaluations or equalize locally determined assessments. Many oversight agencies have a dual role. One role is to advise and assist local valuation offices, and the other role is to measure local valuation performance. These two roles may create a conflict of interest that should be minimized. See Section 13– Considerations for Oversight.

4. STEPS IN RATIO STUDIES

Ratio studies generally involve the following seven steps, however some may occur in a different order or concurrently. Some of the steps, such as the collection and preparation of market data, are continuous processes. These are the steps in a ratio study:

1. Define the purpose, scope, objectives, and limitations of the study
2. Design the study
3. Collection and preparation of market data
4. Sampling
5. Stratification
6. Statistical analysis and evaluation of results
7. Report steps and results of study

Principles

- A ratio study is an iterative process of seven steps; some steps are repeated or done in a different order depending on the purpose of the study.
- Design of the study should reflect the objectives and identify limitations of the study.
- Data collection, sampling, and stratification should be in accordance with acceptable statistical methods.
- Reporting on the methodology, definitions and results is a critical step in the ratio study process.

4.1. DEFINE THE PURPOSE, SCOPE, OBJECTIVES, AND LIMITATIONS OF THE STUDY

The first step in any ratio study is to determine and state clearly the reasons for the study. This crucial step of identifying the purpose of the study determines the specific goals, scope, content, depth, and required flexibility. Some of the reasons for conducting a study include:

- Review existing values/models to determine appraisal priorities
- Calibrate and refine the underlying model
- Validate a model
- Determine market trends
- Determine need for revaluation
- Oversight function – monitor, certify, and validate values. Consider corrective action.
- Equalization (direct or indirect)

4.1.1 Limitations

Users of ratio studies should recognize there may be limitations:

- A ratio study cannot provide perfect information about valuation performance. Lack of sufficient sales, outliers, or overrepresentation of one geographic area or type of property can distort results and make them less reliable.
- Ratio study validity requires that sold and unsold parcels be appraised at the same level and in the same manner. Violation of this condition seriously undermines the validity of both the study and the valuations.
- Findings should be used only in ways that are consistent with the intended use(s) for which the study was designed.
- Ratio study data are subject to statistical sampling errors and other processing (non-sampling) errors that may make them less reliable for informed decision making (Lessler & Kalsbeek, 1992).
- Sales must be screened to eliminate those that don't meet the requirements of arm's-length, market sales (Standard on verification and adjustment of sales, 2020).

- Sales are not "randomly selected" from the population, in the strict technical sense. Depending on the circumstances, sales prices can provide either useful or poor indications of market values (see Section 5.3- Sample Representativeness).
- Value-related characteristics of a sale sample may not represent all the value-related characteristics of the population.

4.2. DESIGN THE STUDY

The ratio study design must consider the quantity of sales data and resources available. It is critical the sales, as a sample, sufficiently represent the population of properties in both characteristics and distribution of values. Although absolute accuracy cannot be ensured, all reasonable, cost-effective steps should be taken to maximize reliability. The analyst should identify the following factors:

- The groups or classes of properties to be included in the study.
- Important legal, physical, and economic characteristics of the properties selected for study.
- The quantity and quality of data available, both overall and within any stratifications
- If value ranges should be created for any continuous type variables such as living area, year of construction, etc.
- The values being tested.
- Sales time period being used and if it is necessary to utilize time adjusted sale price rather than the original sale price.
- In some cases, the analyst can consider the use of independent value estimates as a substitute or supplement to sales. These may be developed using independent mass valuation models. See Appendix A.
- Available resources, such as the number and expertise of staff, computer hardware and software applications, and additional limiting conditions.

4.2.1 Minimum Requirements of a Ratio Study

Although every study does not require the same level of *statistical* detail, each ratio study should include measures of valuation level, valuation uniformity, and confidence intervals showing statistical reliability. Graphs, charts, GIS, or other pictorial representations can be useful tools for showing distributions and patterns in the data. There is no model ratio study design that can serve all jurisdictions or all situations equally well; a ratio study is a function of the quantity and quality of available data and resources. Informed, reasoned judgment and common sense are required in the design of ratio studies.

4.3. COLLECTION AND PREPARATION OF MARKET DATA

Ratio study findings can only be as accurate as the data used in the study. Those involved in collecting, screening, and adjusting sales data should be familiar with real estate conveyance practices in their region.
Data quality checks should follow the guidelines in the Standard on Data Quality (International Association of Assessing Officers (IAAO), 2021), which outlines protocols for accuracy, completeness, currency and consistency of mass valuation datasets.

The legal and physical characteristics of each property used in the ratio study must be the same as when sold. This implies two essential steps. First, the analyst must ascertain whether the property descriptions match (the legal). Second, the analyst must verify if the physical characteristics of the property have changed since the last valuation, and if so, adjustments may be necessary before including the property in a ratio study. Properties with significant differences in these factors should be excluded from the study (Standard on verification and adjustment of sales, 2020).

4.4. SAMPLING

A ratio study is a form of applied statistics because the analyst draws conclusions about the population of properties based on those that have sold during a given period of time. The sales constitute the sample that will be used to draw conclusions or inferences about the population.

It is impossible to determine the accuracy of valuations with absolute certainty. Each sale is only one observation utilized in determining the most probable price (market value) of a property. Each of these observations will have slightly random deviations due to buyer or seller preferences or other unknown issues not identified when confirming the sale. However, when utilized as a group or sample, ratio studies help draw inferences or conclusions about the population from these samples.

It is important to ensure there is a sufficient number of observations (sales) for analysis. With the exception of outliers, all available valid market transactions within the time period should be utilized. When the sample size is inadequate, the analyst should determine if it is possible to expand the time period or geographic area used in the study or use independent valuation proxies in place of sales in order to develop a more robust sample (See Section 9 - Considerations for Small Sample Situations and Appendix A- Independent Value Estimate-based Ratio Studies).

4.4.1 Independent Value Estimate-based Ratio Studies

In specific cases ratio studies can be conducted with estimates of market value that result from independent value estimates. These independent value estimates can be made through fee appraisals or through the deployment of a separate valuation model which has undergone scrutinous testing itself and can be concluded to reflect market value. The use, considerations and principles of independent value estimate-based ratio studies is explained in Appendix A- Independent Value Estimate-based Ratio Studies. Furthermore, the underlying principles in independent value estimate-based ratio studies closely resemble considerations for personal property ratio studies, which is elaborated on in Section 12- Personal Property Ratio Studies.

It is necessary that the use of independent value estimates is reflected in the scope and design of the intended ratio studies. In presenting the results of a ratio study which includes the use of an independent value estimate, it must be clear that the utilized ratios are assessment to model value ratio (AMR) and not assessment to sales price ratio (ASR) (Hermans, McCord, Davis, & Bidanset, 2023).

4.5. STRATIFICATION

Stratification divides all properties within the scope of the study into two or more groups or strata. Stratification facilitates a more detailed picture of valuation performance and can enhance sample representativeness.

Strata should be chosen to be consistent with factors in the mass valuation model and ensure representativeness from the sample. When the purpose of the study is to evaluate valuation quality, flexibility in stratification is essential. The general goal is to identify areas in which the value levels are too high/low or lack uniformity and determine property groups for which additional work may be required. In such cases, it also is highly desirable to stratify on the basis of more than one characteristic simultaneously.

Stratification can help identify differences in level of valuation between property groups. In large jurisdictions, stratification by geographic areas is generally more appropriate for residential properties, while stratification of commercial properties by either geographic area or property subtypes (e.g., office, retail, and warehouse/industrial) can be more effective.

4.6. STATISTICAL ANALYSIS AND EVALUATION OF RESULTS

A properly designed ratio study is a powerful tool for analyzing valuation performance, evaluating mass valuation models, and suggesting strategies for improvement. A ratio study is an iterative process and should be performed throughout the valuation cycle (See Section 5.4- Frequency of Ratio Studies).

A ratio study can identify weaknesses in valuation system performance. Unexpected study results may indicate a need to respecify or recalibrate a valuation model or to reevaluate the data elements used in the valuation process. Measures of valuation level, uniformity, and reliability for the entire jurisdiction, and each group or stratification, should be computed after sales have been screened and ratios have been computed. If appropriate, identified outliers should be removed. Confidence intervals should be applied to validate sample size and determine compliance with standards as well as statistical reliability (see Section 7- Level of Valuation Statistics and Section 8- Uniformity of Valuation Statistics).

4.7. REPORTING

The ratio study report offers clear, data-driven evidence of how valuations are conducted, aids internal decision-making and serves as a communication tool to foster public confidence in the equity of the valuation portion of the tax system. The purpose of a ratio study report is multifaceted, serving several critical functions in property valuation and public administration.

A significant purpose is its role in public relations and transparency. The ratio study report builds public trust by showing that valuations are based on objective market data. By making the process transparent, it addresses concerns from taxpayers, property owners, and stakeholders, demonstrating the property tax system is fair and accountable. Through these multiple purposes, the ratio study report becomes an essential component of effective property tax administration.

As such, transparency in how the study was done is critical. The report should explain the purpose of the study, identify data sources and describe the methodology employed, including sales time frame, data edits, time trends and other price adjustments, outlier trimming methods, statistics used, and reliability and proper use of results. It is also recommended that the report reference a source where interested parties can access data used in the study and potentially replicate study results.

5. SAMPLE SELECTION AND TIMING OF STUDIES

There are several important considerations regarding sample selection and timing of studies which must be considered in the design of a ratio study.

Principles

- Understanding the composition of the population is important in designing the proper study.
- Sample selection and timing of the study should be carefully considered.
- Inadequate samples must be enhanced by expanding the time period used, applying appropriate adjustments for market conditions, or through the use of independent valuation estimates.
- Careful consideration of the representativeness of the “Sample” to the “Population” is paramount to a good study outcome.

5.1. DATA REQUIREMENTS AND AVAILABILITY

Prior to defining study objectives, the jurisdiction must first determine the availability of data which will influence the design of the study. Lack of an adequate sample size may limit the usefulness of the calculated statistics. Since conclusions can be drawn about the valuation of the population of properties based on a sample, it is critical exploratory data analyses are conducted to increase understanding of both sales samples and population datasets. Increased understanding of data quality and characteristics will improve the decision-making process associated with conducting and interpreting ratio study analyses.

5.1.1 Nature of the Population

Understanding the types of properties, market conditions, and composition of the population, in terms of age, size, and value range, are essential to designing the proper study and successfully interpreting the results. Properties that rarely sell (e.g., hospitals) can be considered outliers that would not necessarily be included in a ratio study designed to evaluate performance, however they would be used in a study designed to calibrate the underlying model.

5.1.2 Composition of the Sale Ratio

Values or assessments are generally the numerators in the ratios used in a study. It is important to understand the valuation conditions as this can affect the relationship to the denominator. These conditions include valuation dates, legal requirements, and the time period they remain in effect.

Sale price or proxy, as an indicator of market value, is the denominator in the calculation of the ratio. Specific information about the date, amount, terms, and conditions of each sale is needed for proper analysis.

5.2. PERIOD FROM WHICH SALES ARE DRAWN

Sales used in ratio studies can span multiple years provided there have been no significant changes to property characteristics since the transaction, and sales prices have been adjusted for market conditions. The determination of the number of years utilized in the study is based on providing a representative sample for the analysis.

5.3. SAMPLE REPRESENTATIVENESS

In general, a ratio study is valid to the extent the sample is sufficiently representative of the population, to the extent the distribution of sold properties in the sample reflects the distribution of properties in the population. If sold and unsold properties are valued in the same manner and the sample is otherwise representative, statistics calculated in a sales ratio study can be used to infer valuation performance for unsold properties.

However, if properties that sell are selectively revalued based on their sale prices or not treated the same as the population, and such properties are used in the ratio study, uniformity inferences will not be accurate (valuations appear more uniform than they are). In this situation, measures of valuation level will not be supportable (see

Appendix B- Sales Chasing Detection Techniques). A quality control program should be developed, including checks and audits of the data, to ensure sold and unsold properties are valued at the same level. Holdout samples and other techniques can be used for this purpose and are discussed in the Standard on Automated Valuation Models (International Association of Assessing Officers (IAAO), 2018).

Representativeness is improved when the following occur:

1. Procedures and methodology used to value the population are also used to value the sample. For example, a sample comprised mostly of new construction, first-time sales of improved properties, or new parcels is unlikely to be fully representative of the unsold properties.
2. Recorded property characteristics data for sold property does not differ substantially from that of unsold property. For example, utilizing characteristics only available for the sales, yet not available for the population, such as interior condition or remodeling, would undermine the true representativeness of the sample.
3. Sample properties are not unduly concentrated in certain areas or property types. For example, sales of newly subdivided lots would likely not be representative of the value of land in older, established subdivisions.
4. Sales have been appropriately screened (see Standard on Verification and Adjustment of Sales).

5.3.1 Estimating Performance for Unsold Properties

If unsold properties within a properly specified group are not appraised consistently with sold properties within the same group and according to applicable guidelines, unadjusted sales ratio results cannot be used.

Once it is determined that sales chasing probably has occurred and probably is reducing the validity of ratio study statistical measures of level or uniformity, it is necessary to redo the ratio study to establish valid measures before any other recommendations, such as reappraisal or equalization action, can be made. If feasible, the best approach is to select a sample period that effectively precludes sales chasing. For example, when the lien or appraisal date is January 1, many jurisdictions use sales occurring before that date to make valuation decisions. To test the resulting valuations, it would be appropriate to use sales occurring after January 1 (or after the last date for changing assessments for the year in question), provided such data are time-adjusted (when necessary) backward to match the appraisal date. If sales chasing is not the issue, earlier sales before the reappraisal time period could be used.

5.4. FREQUENCY OF RATIO STUDIES

The purpose of a ratio study dictates how often it should be conducted. Regardless of the revaluation cycle, ratio studies should be conducted at minimum on an annual basis. This frequency enables potential problems to be recognized and corrected before they become serious.

When there is a revaluation, assessors should conduct at least four ratio studies to establish the following:

1. a baseline of current appraisal performance
2. preliminary values so that any major deficiency can be corrected
3. values used in assessment notices sent to taxpayers
4. final values after completion of the first, informal phase of the appeals process

The final study can be used in planning for the following year. In addition, ratio studies can be conducted as needed to evaluate appraisal procedures, investigate a discrimination complaint, or answer a specific question.

6. RATIO STUDY GENERAL STATISTICAL CONCEPTS

Once data has been properly collected, reviewed, assembled, and adjusted, outlier handling and statistical analysis can begin. This process involves the following steps:

1. A ratio should be calculated for each observation in the sample by dividing the value by the sale price.
2. Graphs and exhibits can be developed that show the distribution of the ratios and tests of the hypotheses of normality may be conducted.
3. Exploratory data analysis, including outlier identification and screening should be conducted.
4. Ratio study statistics of both valuation level (Section 7) and uniformity (Section 8) should be calculated.
5. Hypothesis tests should be performed, and reliability measures, such as confidence intervals, should be calculated.

Principles

- Visualizations aid in illustrating general patterns or trends but should not be used to draw formal conclusions.
- Tests of normality should be done to determine if parametric or non-parametric tests should be used.
- Outliers can distort results and should be identified and reviewed or trimmed as required.

6.1. DATA VISUALIZATIONS

Visualizations or exhibits that provide a profile or picture of ratio study data are useful for illustrating general patterns and trends, particularly to non-statisticians. The form of the visualization, as well as the data used (e.g., sales prices, sales ratios, and property characteristics) depends on the purpose of the analysis. Types of visualizations useful in ratio studies include tables, bar charts, histograms, scatterplots, and maps (Gloudemans R. , 1999).
Graphic visualizations can be used to:

- Give a visual representation of sample representativeness
- Illustrate the degree of nonnormality in the distribution of ratios
- Illustrate the overall level of valuation, degree of valuation uniformity, and degree of value related bias (progressivity and regressivity)
- Compare the valuation level or degree of uniformity between strata
- Illustrate potential ratio or value outliers
- Identify specific opportunities to improve mass valuation performance
- Track performance measures over time
- Illustrate market condition (time) trends over time

Graphic visualizations alone should not be used to draw formal conclusions. Graphical representations can be important tools to better understand the nature of the ratio data or to shed light on a particular relationship, however, graphical relationships can also be deceiving and may not consider properly designed hypothesis testing and an accounting of statistical significance in drawing conclusions.

Table 1- Example of Ratio Study Statistical Analysis Data Analyzed

Observation	Assessed Value (AV) ($)	Sale Price (SP) ($)	Ratio (AV/SP)
1	240,000	690,000	0.348
2	144,000	296,250	0.486
3	392,000	787,500	0.498

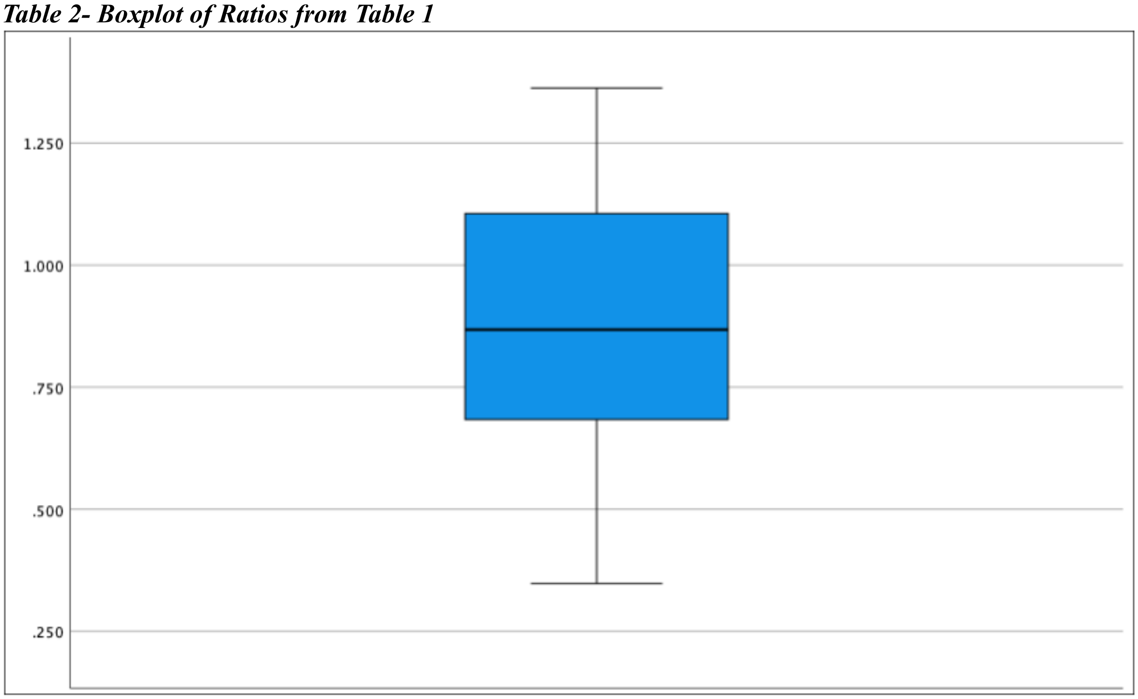
Observation	Assessed Value (AV) ($)	Sale Price (SP) ($)	Ratio (AV/SP)
4	199,200	372,000	0.535
5	340,800	574,500	0.593
6	472,000	795,000	0.594
7	336,000	559,500	0.601
8	284,800	465,000	0.612
9	436,000	693,600	0.629
10	191,200	298,500	0.641
11	225,000	350,000	0.643
12	194,000	295,000	0.658
13	481,600	732,000	0.658
14	338,400	495,000	0.684
15	164,800	237,000	0.695
16	252,800	352,500	0.717
17	945,000	1,289,000	0.733
18	280,000	379,000	0.739
19	407,000	540,000	0.754
20	423,000	560,000	0.755
21	306,800	390,000	0.787
22	236,800	300,000	0.789
23	361,000	450,000	0.802
24	290,400	345,000	0.842
25	235,200	277,500	0.848
26	415,000	485,000	0.856
27	381,000	442,000	0.862
28	380,000	435,000	0.874
29	875,000	998,000	0.877
30	680,000	772,500	0.880
31	329,000	365,000	0.901
32	457,000	487,000	0.938
33	516,000	547,500	0.942
34	295,200	300,000	0.984
35	580,000	580,000	1.000
36	840,000	840,000	1.000
37	640,000	622,500	1.028
38	660,000	637,500	1.035
39	354,000	340,000	1.041
40	800,000	750,000	1.067
41	390,000	352,500	1.106

Observation	Assessed Value (AV) ($)	Sale Price (SP) ($)	Ratio (AV/SP)
42	800,000	705,000	1.135
43	469,000	410,000	1.144
44	1,000,000	859,500	1.163
45	920,000	787,500	1.168
46	527,000	440,000	1.198
47	800,000	648,000	1.235
48	786,000	630,000	1.248
49	496,000	388,500	1.277
50	1,000,000	765,000	1.307
51	320,000	243,750	1.313
52	960,000	720,000	1.333
53	952,000	705,000	1.350
54	327,200	240,000	1.363

Note: Due to rounding, totals may not add to match those on following table, which reports results of statistical analysis of above data.

The following boxplot using the sales ratios contained in Table 1 is an example of data visualization.

Table 2- Boxplot of Ratios from Table 1



6.2. THE NORMAL DISTRIBUTION

Many conventional statistical methods assume the sample data conform to the shape of a bell curve, known as the normal (or Gaussian) distribution. Performance measures based on the mean or standard deviation can be misleading if the study sample does not meet the assumption of normality. As a first step in the analysis, the distribution of sample ratios should be examined to reveal the shape of the data and uncover any unusual features. Although ratio study samples typically do not conform to the normal distribution, statistical tests can be used to evaluate whether there is significant statistical evidence of non-normality. Presence of normal distribution allows for use of parametric tests and provides a more precise interpretation for statistics like Standard Deviation.

6.3. OUTLIER RATIOS

Outlier ratios are very low or high ratios as compared with other ratios in the sample. The validity of ratio study statistics used to make inferences about population parameters could be compromised by the presence of outliers that distort the statistics computed from the sample. One extreme outlier can have a controlling influence over some statistical measures. However, some measures, such as the median ratio, are more resistant to the influence of outliers and trimming may not make a practically significant difference.

Outlier ratios can result from any of the following:

1. an erroneous sale price
2. a non-market sale
3. unusual market variability
4. a mismatch between the property sold and the property valued
5. an error in the value of an individual parcel
6. transcription or data handling errors

In preparing any ratio study, outliers should be:

1. Identified
2. Scrutinized to validate the information and correct errors
3. Trimmed, if necessary

Value outliers have values (e.g. sale prices or estimated values) at the extreme ends of their distribution. In addition to being affected by ratio outliers, statistics such as the weighted mean, price-related differential (PRD) and coefficient of price-related bias (PRB) are also affected by value outliers, even if the ratios on sample properties with atypical values do not appear unusual relative to others. For guidelines on outlier identification and trimming, see Appendix C- Outlier Trimming Guidelines.

6.4. TESTS OF HYPOTHESES

An appropriate test should be used whenever the purpose of a ratio study is implicitly or explicitly to test a hypothesis. A hypothesis is essentially a tentative answer to a question, such as, are residential and commercial properties valued at equal percentages of market value? A test is a statistical means of deciding whether the answer "yes" to such a question can be rejected at a given level of confidence.

Tests are available to determine whether the:

- valuation level of a stratum fails to meet an established standard
- meaningful differences exist in the level of value between two or more strata
- high-value properties are valued at a different percentage of market value than low-value properties

Table 3- Tests of Hypotheses

Null Hypothesis	Nonparametric Test	Parametric Test
1. Ratios are normally distributed	Shapiro-Wilk W test D'Agostino-Pearson K2 test Anderson-Darling A2 test	N/A
2. The level of valuation meets legal requirements.	Binomial test (Median Sign Test)	t-test
3. Two property groups are appraised at equal percentages of market value.	Mann-Whitney Test Mood's Median Test	Two-sample t-test
4. Three or more property groups are appraised at equal percentages of market value.	Kruskal-Wallis Test	ANOVA
5. Low or high value properties are appraised at equal percentages of market value.	Spearman Rank Test, GINI-based measures; Clapp's Two-Stage Least Squares Approach	PRB
6. Sold and unsold parcels treated equally.	Mann-Whitney Test Mood's Median Test	Two-sample t-test

Other appropriate tests are discussed in (Gloudemans R. , 1999), (Property appraisal and assessment administration, 1990), and (Improving real property assessment: A reference manual, 1978).

6.5. PARAMETRIC AND DISTRIBUTION-FREE (NON-PARAMETRIC) STATISTICS

For every problem that might be solved by using statistics, there is usually more than one measure or test. These measures and tests can be divided into two broad categories: parametric and distribution-free (nonparametric). Parametric statistics assume the population data conform to a known family of probability distributions (such as the normal distribution).

When measures like the mean, weighted mean, and standard deviation are used in the context of a normal distribution, they tend to be more meaningful. Distribution-free statistics make less restrictive assumptions and do not require knowledge about the nature of the underlying population distribution. Given similar distribution of ratios in the underlying populations, distribution-free tests, such as the Mann-Whitney test, can determine the likelihood that the valuation level of property groups differ. Distribution-free measures include the median and the COD.

7. LEVEL OF VALUATION STATISTICS

The first of two primary aspects of valuation accuracy is the level of valuation. Valuation level refers to the overall, or typical, ratio at which properties are valued relative to market proxies.

Estimates of valuation level are based on measures of central tendency. They should be calculated for each stratum and for aggregations of strata as may be appropriate. Several common measures of valuation level (central tendency) can be calculated in ratio studies, including the median ratio, mean ratio, and weighted mean ratio.

When one of these measures is calculated on the data in a sample, the result is a point estimate, which is accurate for the sample but is only an indicator of the level of valuation in the population. Therefore, confidence intervals should be calculated to determine the reliability of these point estimates as predictors of the overall level of valuation of the population.

Particularly in oversight situations, it is important that a finding of noncompliance with valuation level standards **should not** be determined based on sample point estimates alone, except as provided in 13.6.1.5. This requires the use of confidence intervals or statistical hypothesis tests.

Principles

- Measures of central tendencies such as median, mean and weighted mean should be calculated to estimate level of valuation.
- Measures computed from samples are point estimates, serve as indicators, and require confidence intervals to provide measures of statistical significance.
- Different tests of level have different uses and applicability.

7.1. MEDIAN

The median ratio is the middle ratio when the ratios are arrayed in order of magnitude (ascending or descending). If there is an even number of ratios, the median is the average of the two middle ratios.
The median divides the data into two equal parts and is less affected by extreme ratios than other measures of central tendency. Because of these properties, the median is the generally preferred measure of central tendency for

Median
Middle value of the dataset
0.82, 0.92, 0.99, 1.05, 1.15
= 0.99

evaluating overall valuation level, determining revaluation priorities, or evaluating the need for a revaluation.

7.2. MEAN

The mean ratio is calculated by summing the ratios and dividing by the number of ratios.
The mean is affected more by extreme ratios than the median. In a distribution skewed to the right (typical of ratio

Mean

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

0.82, 0.92, 0.99, 1.05, 1.15
= 0.986

study data), the mean is greater than the median.

7.3. WEIGHTED MEAN

The weighted mean ratio is the value-weighted average of the ratios in which the weights are proportional to the sales prices. The weighted mean also is called the aggregate ratio.
The weighted mean can be calculated by (1) summing the values, (2) summing the sales prices, and (3) dividing the first result by the second.

Weighted Mean Ratio
The Weighted mean ratio is the weighted average value

where x = values and y = sale prices.

Parcel	Value	Sale Price
1	375,000	325,000
2	520,000	480,000
3	250,000	240,000
4	400,000	438,000
5	700,000	850,000
Sum	2,245,000	2,333,000

$$Wtd\ Mean = \frac{\sum x}{\sum y}$$

$$Wtd\ Mean = \frac{2{,}245{,}000}{2{,}333{,}000} = 0.962$$

The weighted mean gives equal weight to each unit of value (e.g. dollar) in the sample, whereas the median and mean give equal weight to each property. The weighted mean is an important statistic and is also used in computing the PRD, a measure of uniformity between high- and low-value properties. The weighted mean is affected more by extreme ratios on high priced properties.

7.4. CONTRASTING MEASURES OF VALUATION LEVEL

The median is the preferred measure of central tendency for evaluating appraisal performance because it gives equal weight to each ratio and is unaffected by extreme ratios. Although the mean ratio also gives equal weight to each ratio, it can be affected appreciably by extreme ratios and can be relied upon only if outliers have been appropriately trimmed. The weighted mean is also affected by extreme ratios. It is the generally recommended measure of central tendency for indirect equalization because it gives equal weight to each unit of value (dollar).

7.5. MEASURES OF RELIABILITY

Reliability, in a statistical sense, indicates the degree of confidence that can be placed in a statistic calculated from a sample which is then being used as an estimate for an unknown population. Sampling error always exists and diminishes as sample size increases or variability decreases. Using measures of reliability provides critical information to determine if a sample point estimate is truly in compliance.

7.5.1 Confidence Intervals

Reliability is measured with a confidence interval which brackets a sample's point estimate measure with two numbers, an upper and lower limit. These limits utilize a specified degree of confidence (the most common in valuation being 90% and 95%) which determines the upper and lower limits of the true measure of central tendency for the population. If the point estimate is out of compliance, but either the upper or lower limit overlaps the expected parameters, then the analyst cannot reliably say the test has failed; see section 13.6.1.4, Table 12.

Confidence intervals explicitly consider errors inherent in the sampling process. In general, larger and more uniform (i.e. less disperse) samples produce narrower confidence intervals while smaller and less uniform (i.e. more disperse) samples produce wider confidence intervals. These confidence intervals indicate whether the desired degree of confidence has been achieved regarding a given parameter or range. This is important in ratio studies because the reliance on samples means the true measures of the entire population are unknown and care should be taken to ensure any given statistic truly meets or fails to meet standards.

The analyst should not tolerate measures of central tendency that fail to meet goals or standards whenever measures of reliability are wide due to small samples, poor uniformity of values, or both. Such cases require either additional data for proper analysis or alternative action if poor uniformity is the cause. Such alternative action might include revaluation, trending of strata, or respecifying or recalibrating mass valuation models.

See Appendix 20- 4 in (Property appraisal and assessment administration, 1990) and Appendix D for guidelines on calculating small-sample confidence intervals. Procedures such as bootstrapping (Efron & Tibshirani, 1993), enable the development of confidence interval estimates for any statistic of interest, including measures of level and uniformity.

Standard formulas for calculating confidence intervals around measures of central tendency and the COD can be found in Appendix D.

8. UNIFORMITY OF VALUATION STATISTICS

In addition to the valuation level statistics discussed in Section 7- Level of Valuation Statistics, uniformity between groups of property is an equally important and critical aspect of mass valuation and ratio studies. When evaluating uniformity, there are two aspects: horizontal equity, which looks at how uniform ratios are between groups or stratum, and vertical equity, which looks at how uniform ratios are as value increases.

Principles

- Horizontal and vertical equity measures should be computed and examined to determine uniformity and compliance with standards.
- Conducting multiple tests may be required to ensure compliance.
- Extremely low variability should be examined to detect possible sales chasing or model overfitting.
- Different classes and stratum should achieve similar measurement results.

8.1. MEASURES OF VARIABILITY

Measures of dispersion or variability of the ratio distribution relate to the uniformity of the ratios. In general, the smaller the dispersion, the better the uniformity. However, extremely low measures can signal the following:
Typical acceptable causes:

- extremely homogeneous properties
- very stable markets

Unacceptable causes:

- poor sample representativeness (Section 9)
- sales chasing and model overfitting (Appendix B)
- ratio over-trimming (Appendix C)

It is important to note that as market activity changes or as the complexity of properties increases, the measures of variability will typically also increase, even though valuation procedures may be equally valid. Measures of dispersion should be calculated overall and for each stratum in the study. By evaluating across stratum, the presence or absence of horizontal equity can be determined.

8.1.1 Coefficient of Dispersion

The Coefficient of Dispersion (COD) is generally considered the most useful measure of variability or uniformity in property valuation. The COD measures the average absolute percentage deviation of the ratios from the median ratio and is calculated by the following steps:

1. Calculate the overall sample median ratio.
2. Subtract the median from each individual ratio.
3. Determine the absolute value of the calculated differences.
4. Sum the absolute differences.
5. Divide by the number of ratios to obtain the average absolute deviation.
6. Divide by the overall sample median ratio.
7. Multiply by 100.

A desirable feature of the COD is that its interpretation does not depend on the assumption that the ratios are normally distributed. In general, more than half the ratios fall within one COD of the median.

Note that the typical percentage error is not the COD but is expressed by the median absolute percentage error statistic. Also, it is the interquartile range, not the COD, that brackets the middle 50 percent of the assessment ratios.

8.1.2 Standard Deviation and Coefficient of Variation

The standard deviation and coefficient of variation (COV) are common measures of dispersion and can be a powerful measure of appraisal uniformity. These statistics are calculated using the following steps:

1. Determine the overall sample mean ratio.
2. Subtract the mean from each individual ratio.
3. Square the resulting differences.
4. Sum the squared differences.
5. Divide by the number of ratios less one to obtain the variance of the ratios.
6. Compute the square root of the variance to obtain the standard deviation.
7. To obtain the Coefficient of Variation, divide the standard deviation by the mean ratio.

For reference, in a normal distribution, the COV is approximately 1.25 times the COD.

8.1.3 Other Measures of Variability

Other useful measures of variability applicable to ratio studies include:

- range: A measurement of dispersion. Higher is associated with more variability
- percentiles (e.g. quintiles, quartiles, and deciles): Dividing the data set into separate strata for analysis.
- interquartile range: A measure of dispersion of the middle 50% of the data.
- median absolute deviation (MAD): Measure of variability given by the median of absolute deviation from the median.
- mean absolute percentage error (MAPE): Measure of accuracy of predictive models
- median absolute percentage error (MdAPE): Measure of accuracy of predictive models which is less susceptible to outliers due to its use of the median.
- coefficient of concentration: Relative measure of variation in terms of the mean.
- coefficient of variation (COV): Measure variation given by the ratio of standard deviation to the mean.

8.2. VERTICAL EQUITY

Vertical equity evaluates consistency in valuation levels across the range of market values. When there are systematic differences, this is termed vertical inequity. Vertical equity measures inherently provide an indication of the relationship between the ratio and a market value proxy.

When low-value properties are valued at greater percentages of market value than high-value properties, assessment regressivity is indicated. When low-value properties are appraised at smaller percentages of market value than high-value properties, assessment progressivity is the result. Valuations made for tax purposes should be neither significantly regressive nor progressive.

As part of the valuation process, practitioners should ensure that valuation levels are consistent between various property characteristics such as geographic areas, size groups, age groups, etc. If valuation levels are consistent between these various property groups, then they are likely consistent between value ranges. If inequities are found among property groups during the valuation process, they should be addressed before undertaking a formal vertical equity analysis.

Once values are finalized, agencies should conduct a formal vertical equity analysis to ensure IAAO or relevant jurisdiction standards are achieved.

Various vertical equity measures and analyses are available. Each method has different strengths and weaknesses and may provide different insights and results. For purposes of compliance with IAAO standards, primary emphasis should be placed on the Vertical Equity Indicator (VEI). Supplemental measures provide additional insight and can potentially confirm the finding of the VEI.

Three of these traditional and/or effective measures are the Price Related Bias (PRB), Modified Kakwani Index (MKI), and the Price Related Differential (PRD).

8.2.1 Vertical Equity Measures

The following table summarizes the features of each of the vertical equity measures noted above. Some measures and tests can be useful in exposing and diagnosing vertical equity patterns, however they stop short of indicating whether vertical inequity rises to the point of being provably unacceptable. The Vertical Equity Indicator (VEI) is a comprehensive test that fills this void.

Table 4 – Comparison of Vertical Equity Measures

Measure	Quantifiable	Compliance Testing	Visualization
VEI[1]	✓	✓	✓
PRB[2]	✓	✓	✓
MKI[3]	✓	✓	X
PRD[4]	X	X	X

1: Vertical Equity Indicator (VEI)
2: Price Related Bias (PRB)
3: Modified Kakwani Index (MKI)
4: Price Related Differential (PRD)

Table 5 – Descriptions of Vertical Equity Measures

Measure	Vertical Equity Measures
VEI	Primary measure. Used to determine compliance with IAAO standard.
PRB	Measures percentage change in assessment ratios each time values double or are halved. PRBs of ±.05 tend to indicate good equity.
MKI	A statistic measuring how the share of the cumulative assessed values compares with the cumulative share of sales prices.
PRD	The ratio of the value-weighted mean to the mean ratio. Ratios of 0.98 to 1.03 can indicate good equity, although the measure is highly influenced by value and ratio outliers.

Details to calculate the VEI are included in Appendix E. Further Information on the remaining measures and additional vertical equity measures are included in Appendix F. Vertical Equity testing should be completed for large geographic areas (e.g., market area, county, municipality), and across major property categories or tax classes (e.g., residential, commercial, vacant land). Testing at the sub-category level (e.g. neighborhood) may be considered if sufficient sales are available for analysis. Testing for vertical inequity over a stratum with a relatively narrow range of market value is usually less meaningful.

9. CONSIDERATIONS FOR SMALL SAMPLE SITUATIONS

There is a general relationship between statistical reliability and the number of observations in a sample. The larger the sample size, the greater the reliability. Even large jurisdictions will face small sample situations for certain property types, such as gas stations, where there are limited numbers in the population.

Principles

- The usefulness of the statistics is a function of sample size; smaller samples provide less meaningful statistics.
- Point estimates alone are inadequate for determining compliance for a small sample; confidence intervals must be used to reflect sample error.
- Sample size can be improved by extending the time period, re-stratification, use of independent estimates, or validating previously rejected sales.

9.1. ADEQUACY OF A GIVEN SAMPLE SIZE

The adequacy of a given sample size can be evaluated by computing measures of reliability. Measures such as confidence intervals are wider given smaller sample sizes. This makes drawing statistically valid conclusions about applicability of sample results to intended populations more questionable. The two ways to address this problem are to: (1) accept less precision (ie: lower the degree of confidence being used from say 95% to 90% or 80%, etc.) or (2) increase the sample size by including longer sampling periods, combining strata, or adding independently developed valuations as proxy sales (see Section 9.3 for further explanation of expansion methods).

9.2. REQUIRED SAMPLE SIZE

The minimum sample size is relative to the representativeness of the sample. If the sample consists of mostly new construction or new parcels, the applicability of the results to the population could be minimal. To increase the reliability of the statistical measures and the guidance they provide, the sample size should be increased to ensure the market data being used is representative of the population as a whole.

9.3. REMEDIES FOR INADEQUATE SAMPLES

Inadequate sample sizes are typically indicated by unacceptably wide confidence intervals (see Section 9.1). The following alternatives should be considered:

1. ***Re-stratification*.** If levels of appraisal are similar or properties are homogenous, broader strata containing larger samples can be created by combining existing strata or by stratifying on a different basis.
2. ***Extending the period from which sales are drawn*.** This is often the most practical and effective approach. Sales from prior years can be used; however, adjusting the sale price for time may be necessary and significant property characteristics must not change.
3. ***Enlarging the sample by validating previously rejected sales*.** Sales previously excluded from the analysis, because it was not administratively expedient to confirm them or to make adjustments to the sale price, can be reevaluated.
4. Imputing appraisal performance. Ratio study statistics for strata with no or few sales can sometimes be imputed from the results obtained for similar strata provided the procedures and techniques used to appraise properties are consistent. If a property type were used to impute performance for a dissimilar property type solely due to sample size, then conclusions would be unreliable or unsupportable.

5. ***Conducting independent value estimate-based ratio studies.*** Market value estimates are developed independently from the assessor's values via independent automated valuation models or single-property appraisals. This approach minimizes bias and allows greater control over sample selection, improving representativeness through random or stratified sampling in underrepresented areas or property types.

10. PRESENTATION OF FINDINGS, DOCUMENTATION, AND TRAINING

The findings of a ratio study should be sufficiently detailed and documented to meet the needs of the users of the study. Documentation for internal ratio studies can be less detailed than for reports prepared for external uses (See Section 4.7- Reporting).

Principles

- Documentation should be created to support and adequately explain the study in accordance with academic and IAAO standards.
- Training and education should be provided and focus on interpretation and use of results.

10.1. DOCUMENTATION FOR PUBLISHED STUDIES

The following documentation should be provided in conjunction with any published ratio study.

10.1.1 Narrative

A ratio study should be accompanied by a brief narrative describing the purpose and the methods used. This information can be incorporated in the report of the findings or be contained in a separate memorandum. The narrative should contain the statistics presented and outline the major procedural steps in completing the study and should also describe any rules for eliminating sales or extreme ratios, and acknowledge any significant limitations in the data and include information used to determine that the ratio study sample is representative of the population.

10.1.2 Exhibits

The body of the ratio study report should include for each stratum the statistical results intended to be used for decision-making purposes. Reports should contain the following written or graphical information:

- date and tax year of the appraisals being evaluated
- number of parcels in each stratum
- number of sales
- number of sales trimmed from the study
- measures of central tendency (valuation level)
- measures of uniformity (variability) and price- related biases
- confidence interval (measures of reliability)
- summary of adjustments made to sales prices

10.1.3 Analyses and Conclusions

An objective statement of the results of the ratio study should be prepared. If the study is one in a series, a comparison of the results with those of previous studies can be helpful.

10.2. INTERNAL DOCUMENTATION

In addition to information for published studies, there should be internal documentation. Operating Procedures should include:

- General explanation of the design of the study. This explanation should be updated whenever procedures are changed.
- Documentation of software applications.
- An explanation of how to execute the study or run the software.

10.3. TRAINING AND EDUCATION

The understanding of ratio studies can be improved through education and training. Seminars or workshops should be conducted to explain how to interpret reports, how ratio studies can be used to improve appraisal performance, and how the results will be used in-house.

11. RATIO STUDY STANDARDS

Each property valuation authority and oversight agency should have ratio study performance standards. Local standards should be consistent with state, provincial, or national standards. The standards summarized in Table 7 and Table 8 are suggested for jurisdictions in which current market value is the legal basis for valuation.
In general, when state, provincial, or national standards are not met, revaluation or other corrective measures should be taken, or equalization procedures can be imposed. When an oversight agency orders such actions, the burden of proof should be on the oversight agency to show that the standards have not been achieved.
All standards recommended in this section are predicated on the assumption that steps have been taken to maximize representativeness and validity in the underlying ratio study.

Principles

- Ratio studies must employ measures of statistical significance—such as confidence intervals and hypothesis testing—to evaluate whether appraisal level and uniformity meet established performance standards.
- Some tolerance in evaluating mass valuation performance is acceptable and necessary. Failure to meet tolerance ranges indicates need for remediation.
- Extraordinary circumstances—such as natural disasters or emerging markets—may increase uncertainty, but do not justify the abandonment of fundamental ratio study standards.

11.1. LEVEL OF VALUATION

In analyzing valuation level, ratio studies attempt to measure statistically how close valuations are to market value (or to a required statutory constraint that can be expressed as a percentage of market value) on an overall basis. While the target level of valuation should be 1.00, a valuation level between 0.90 and 1.10 is considered acceptable for any class of property.

By themselves, point estimates for a measure of central tendency provide only an indication, not proof, of whether the level for the population of properties fails to meet the appropriate goal. Statistical tests or an equivalent confidence interval should be used to determine whether it can be reasonably concluded that the valuation level for the population differs from the established goal.

A decision by an oversight agency to take some action (direct equalization, indirect equalization, revaluation) can have serious consequences for taxpayers, taxing jurisdictions, and other affected parties. This decision should not be made without a high degree of certainty that the action is warranted. Conversely, a decision not to take action when action is needed can have equally serious consequences.

11.1.1Purpose of Level-of-Valuation Standard

Jurisdictions that follow the IAAO recommendation of annual revaluations (Standard on property tax policy, 2020) and (Standard on mass appraisal of real property, 2025) and comply with (Uniform standards of professional appraisal practice (USPAP)) should be able to develop mass appraisal models that maintain an overall ratio level of or very near 100 percent. However, the valuation entity may be compelled to follow revaluation cycles defined by a legal authority or public policy that can extend beyond one year. During extended cycles, the influence of inflation or deflation can shift the overall ratio.

The purpose of a performance standard that allows reasonable variation from 100 percent of market value is to recognize the limiting conditions that may constrain the degree of accuracy that is possible and cost-effective within a valuation entity.

11.1.2 Confidence Intervals in Conjunction with Performance Standards

The purpose of confidence intervals and similar statistical tests is to determine whether it can be reasonably concluded that the appraisal level differs from the established performance standard range in a particular instance. A conclusion of noncompliance requires a high degree of confidence; thus, a 90 percent (two-tailed) or 95 percent (one-tailed) confidence level should be used, except for small or highly variable samples. The demonstration ratio study report in Table 6 presents 90% two-tailed confidence interval estimates for the mean, median, weighted mean ratio, and COD using the data from Table 1 in Section 6.1.

Table 4- Demonstration Ratio Study Report

Statistic	Result
Number of observations in sample	54
Total value	$26,127,200
Total sale price	$29,020,600
Average value	$483,837
Average sale price	$537,418
Mean ratio	0.894
90% mean two-tailed confidence interval	(0.835, 0.953)
Median ratio	0.868
90% median two-tailed confidence interval	(0.787, 0.984)
Weighted mean ratio	0.900
90% weighted mean two-tailed confidence interval	(0.836, 0.965)
Coefficient of Dispersion (COD)	24.65%
Coefficient of Dispersion (COD) 90% confidence interval	(21.878%, 28.618%)
Vertical Equity Indicator (VEI)	48.53%
Vertical Equity Indicator Significance	17.57%
Normal distribution of ratios (0.05 level of significance)	Reject
Additional Vertical Equity Measures	
Statistic	**Result**
Price-related bias (PRB) coefficient (t-value)	.1871 (3.113)
Price-related bias 95% confidence interval	(0.0665, 0.3078)
Price Related Differential (PRD)	0.993
Date of Analysis	01/01/2023
Category or class being analyzed	Residential

11.2. VALUATION UNIFORMITY

The purpose of this section is to establish performance standards for valuation uniformity, including measures of variability such as the coefficient of dispersion and related statistics as described in Section 8. These standards are intended to provide guidance in determining when observed variability indicates acceptable appraisal performance and when corrective action may be warranted.

Assuming the existence of an adequate and sufficiently representative sample, if the uniformity of valuations is unacceptable, model recalibration and/or revaluation should be undertaken. It is important to recognize that the COD is a point estimate and an indicator, but not proof of non-compliance. Proof can be provided by recognized statistical tests, including confidence intervals (Bonett & Seier, 2006) or comparing the computed COD against the maximum permissible sample COD that would be accepted for a given standard as described in Table 7, which limits are derived using tests of statistical significance. See Appendix D6.

In unusually homogeneous strata, low CODs can be anticipated. In all other cases, CODs less than what is specified in Table 7 should be considered suspect and possibly indicative of non-representative samples, model overfitting, or the selective revaluation of sold parcels.

11.2.1 Uniformity among Strata

- Although the goal is to achieve an overall level of valuation equal to 100 percent of the legal requirement, ensuring uniformity in valuation levels among strata also is important. The level of valuation for each stratum (neighborhood, age group, market areas, and the like) should be within 5 percent of the overall level of valuation (not 5 percentage points but 5 percent). This test should be applied only to strata subject to compliance testing. Consider an example where the overall median ratio for a residential market area is 1.01 with a 90% confidence interval of .98 to 1.05, each neighborhood within that market area should have a level of valuation within 5% of the overall level of valuation for the market area. It can be concluded that this standard has been met if 90 percent (two-tailed) confidence intervals about the chosen measures of central tendency for each of the strata fall within 5% of the overall median for the market area. Using the above example, if a neighborhood had a median ratio 0.85 with a 90% confidence interval of 0.81 – 0.94, it would be in compliance because the percent change is 4.08% which is less than 5% different when the upper confidence limit in the neighborhood is compared to the lower confidence limit around the overall median.

$$Percent\ Change = \left(\frac{0.98 - 0.94}{0.98}\right) \times 100 = 4.08\%\ rounded$$

In addition, classes or categories should not have valuation levels provably differing from each other by more than 10% - see 13.6.1.2 .

Definitions of percentage points and percent are as follows:

- Percentage Points – A percentage point is the arithmetic difference between two percentages. For example, if the median ratio changes from 90 to 95, it has increased by 5 percentage points: 95 – 90 = 5.

- Percent – Measures the relative change between two values. In this same example of the median ratio starting at 90 and then changing to 95, the percent increase is:

$$Percent\ Change = \left(\frac{New\ Value - Old\ Value}{Old\ Value}\right) \times 100$$

$$Percent\ Change = \left(\frac{95 - 90}{90}\right) \times 100 = 5.56\%\ rounded$$

11.2.2 Uniformity within Strata

In sales ratio studies, uniformity within strata, or horizontal equity, is determined using measures of variability. The Coefficient of Dispersion (COD) is the recommended measure to determine horizontal equity. The lower the measure, the better the uniformity of estimated sale prices or assessed values. As markets become more complex or market activity changes, measures of variability typically increase.

Table 7 outlines the COD Standards by general property class accounting for differences in market profile and activity. Each property category has a general range for non-segmented studies of the entire population. When the study involves a segmented portion of the population or a population with unique characteristics or size, a modified range is provided. For example, a study of the entire residential population would have a typical COD range from 5.0 to 15.0. However, when the study is focused on newer or more homogeneous segments of the population, the typical COD range is 4.0 to 10.0. A segment of more rural, seasonally occupied housing or manufactured homes would typically produce a COD range of 10.0 to 20.0. Not all unique segments or population profiles found in practice are listed. When not specified, additional segmented studies should generally use the overall range for the corresponding property type category.

In unusually homogeneous strata, low CODs can be anticipated. In other cases, CODs less than specified in Table 7 should be considered suspect and possibly indicative of nonrepresentative samples or selective reappraisal of selling parcels. On the other hand, CODs within the acceptable range found in Table 7 do not automatically indicate that selective reappraisal is not an issue. See Appendix B.

Table 7- Ratio Study Uniformity Standards indicating acceptable general quality*

Property Category	Jurisdiction Size/Profile/Market Activity	COD Range
Residential Improved		**5.0 to 15.0**
	• Newer or more homogeneous market areas such as condominiums	4.0 to 10.0
	• Rural, seasonally occupied housing, or manufactured homes	10.0 to 20.0
Multi-Unit Apartments		**7.0 to 20.0**
	• Rural jurisdictions	10.0 to 25.0
Commercial, Industrial, Retail, Warehouse, etc.		**10.0 to 25.0**
	• Commercial/Industrial Condominiums	5.0 to 15.0
	• Less active markets with small sample size	10.0 to 30.0
Vacant Land		**10.0 to 20.0**
	• Less active markets with small sample size	10.0 to 25.0
	• Rural jurisdictions	10.0 to 30.0
Agricultural land		**10.0 to 25.0**

These types of property are provided for guidance only and may not represent jurisdictional requirements.

**Valuation level for each type of property shown should be between 0.90 and 1.10, unless stricter local standards are required.*

**The acceptable IAAO standard range for the VEI is –10% to 10%*

11.2.3 Uniformity Standards for Other Property
Target CODs for special-purpose real property and personal property should reflect the nature of the properties involved, market conditions, and the availability of reliable market indicators. Special-purpose properties tend to sell infrequently and are less likely to transact under arms-length conditions compared to more common property types. As a result, their sale prices often exhibit substantial variability. Calculated CODs for these properties may be heavily influenced by sampling error, and resulting confidence intervals may be wide, limiting the reliability of the measure. Given these limitations, assessors should rely on market expertise and judgment when evaluating uniformity.

11.2.4 Uniformity Standards in Blighted, Distressed, and Emerging Market Areas
COD standards in blighted, distressed or emerging (synonymously, "developing") market areas differ from other market areas due to either the deterioration of or progression to a properly functioning real estate market. Blighted or distressed markets may correspond to the decline stage in a neighborhood's life cycle. The decline stage of a neighborhood's life cycle may reflect diminishing demand or desirability, general property use change, or even renewal, rebuilding, or restoration (Fundamentals of Mass Appraisal Text Book, page 11, 2011). However, emerging markets are much broader, cover a much larger geographical area and generally refer to the maturing nature of markets and infrastructure.

A blighted property can be defined as land or building that has fallen into a significant state of disrepair, making it hazardous, unattractive, or uninhabitable. Common indicators of a blighted property include structural defects or dilapidation, abandonment, and unsafe living conditions.

A market area, neighborhood, or region may be defined as blighted if enough blighted properties exist to alter the functioning of the market due to declining economic conditions and social issues.
Blighted areas differ from distressed areas since properties do not meet the definition of blight, and any changes in the market behavior are more related to financial difficulty leading to foreclosures, short sales, or tax delinquencies. Sales in both a blighted area or distressed area may be limited and not meet the conditions of an open market transaction, making sale prices less reliable indicators of market value that require a different standard to measure horizontal equity.

Emerging real estate markets for purposes of this standard, refers to a market environment where the formal practices of buying, selling, and valuing property are not fully developed.

In some cases where private ownership is well established, institutional systems handling market activity, such as registries, brokerage practices, appraisal standards, or legal oversight, may not be fully established or consistently applied. Additionally, buyers and sellers may be less knowledgeable about the market when agreeing to a sale price. A developing or emerging market may be characterized by one or more of the following:

- There is a lack of trustworthy or accessible data on past sales, making it difficult for buyers and sellers to determine fair market value
- Sale prices are often underreported and/or inaccurately recorded
- True value of properties is often circulated informally through personal networks rather than institutional systems
- Property rights are unevenly enforced
- Land titling systems are incomplete or contested
- Professional services such as real estate agencies, appraisal firms or legal specialists lack adequate industry standards or consistent regulations
- Trust in market mechanisms is low

- Transactions rely heavily on informal relationships, local knowledge, and personal negotiation rather than standardized processes
- Slow emergence of informed market participants
- Arm's Length transactions will have more dispersion in their sale ratios because the market is not as defined

Rather than being defined by rapid growth in activity or prices, a developing real estate market in this context is marked by the gradual emergence of informed market participants, the establishment of trust in formal processes, and the slow building of transparency and professional infrastructure necessary for a stable, efficient marketplace. Determination of areas as blighted, distressed, or emerging should be done, at minimum, on an annual basis to consider if such conditions still exist.

Table 5- COD Standards for Blighted, Distressed, and Emerging Market Areas

General Property Class	Jurisdiction Size/Profile/Market Activity	COD Range
Residential improved (as defined in Table 7)		
	Distressed or blighted market areas	10.0 to 25.0
	Emerging market areas	10.0 to 35.0
Multi-Family		
	Distressed or blighted market areas	10.0 to 30.0
	Emerging market areas	10.0 to 35.0
Commercial, Industrial, Retail, Warehouse, etc.		
	Distressed or blighted market areas	10.0 to 35.0
	Emerging market areas	10.0 to 40.0
Vacant Land		
	Distressed or blighted market areas	10.0 to 30.0
	Emerging market areas	10.0 to 45.0
Agricultural land		10.0 to 45.0

11.2.5 Very Low CODs

Very low CODs may occur in highly homogeneous markets where property characteristics and market conditions are unusually uniform. However, CODs that fall below the minimum values, shown in Table 7 and Table 8, for the applicable property type and market should be treated as a diagnostic warning. Such results may indicate issues such as sales chasing (selective reappraisal of sold properties), non-representative sales samples, or model overfitting. When unusually low CODs are observed, practitioners should review model specification, sales screening procedures, and other measures of equity before concluding that the results reflect reliable valuation performance. Tests to detect sales chasing should be considered. See Appendix B.

11.2.6 Vertical Equity

In addition to general measures of uniformity such as the coefficient of dispersion, vertical equity measures are used to evaluate whether the level of valuation is applied consistently across the range of property values.
An analysis of vertical equity provides an indication of whether properties are equitable in relation to an indicator of market value across the range of a market value proxy. The preferred measure of vertical equity is the Vertical Equity Indicator (VEI). The acceptable IAAO standard range for the VEI is –10% to 10%. This means that the median ratio for the lowest-valued property group should be within 10%, adjusted for the overall median, of the median ratio for the highest-valued property group. A VEI that is statistically significantly outside of this range does

not meet the vertical equity standard. A statistical testing strategy was suggested for this in Section 8.2.1 which uses a quantity called the VEI Significance.

11.3. NATURAL OR OTHER DISASTERS AND RATIO STUDY STANDARDS

All factors listed below should be considered when ratio study standards are being applied to study results from areas substantially affected by disasters. Such consideration should not result in unwarranted relaxation of applicable standards. When faced with such situations, valuers must use informed, reasoned judgment and common sense to produce a sufficiently reliable ratio study, based upon the best information available.

- Natural disasters such as earthquakes, floods, tornadoes, and hurricanes can have a substantial impact on the interpretation and use of ratio studies.
- Other disasters such as ground water contamination or the failure of infrastructure could also impact the interpretation and use of ratio studies
- Properties impacted by disasters may require adjustment to the valuation approach.
- If an adjustment is warranted, it should be based on the level of impact and interpretation from the market.
- A number of unreliable sample properties may need to be excluded, and sample sizes may be unavoidably reduced.

12. PERSONAL PROPERTY RATIO STUDIES

Personal property ratio studies can be done by those primarily responsible for valuation of such property to determine the quality of personal property valuations.

Most personal property ratio studies performed by oversight agencies are performed for equalization purposes and hence sample must be designed accordingly. For direct equalization the level of valuation for property classes or strata subject to such equalization is the primary area of interest and for indirect equalization it is the overall value of the population and the disproportionate influences of high value properties.

Horizontal equity requires similar levels of appraisal between real and personal property. Sales data for personal property are difficult to obtain and analyze because markets for personal property are generally less visible to follow than real property markets. Therefore, performance reviews and appraisal ratio studies should be used in place of sales ratio studies to determine the quality of appraisal of personal property. The performance review does not quantify assessment conditions but can determine general assessment quality. The appraisal ratio study can be used to determine the level and uniformity of assessment for personal property.

Indirect equalization in particular requires overall estimation of value based on validation of property owner filings and/or estimated values by assessing officers for property owners that did not report accurate costs. The validation process is most typically accomplished by conducting personal property audits and/or physical inspections.

Principles

- Audits and/ or physical inspections are essential for validating the values reported by or estimated for the property owner.
- Performance reviews aid in evaluating assessment performance by providing alternate methods of evaluating level and uniformity of personal property valuation.
- Personal Property escaping assessment must be identified and factored into performance reviews to calculate a corrected level of valuation that reflects both reported and unreported property. Similarly, identifying and removing property that is no longer a part of the assessment roll is also needed to calculate the correct level of valuation.

12.1. THE PERFORMANCE REVIEW

The performance review is an empirical study that evaluates the assessment method used and the ability of the jurisdiction to meet its legal requirement in the assessment of personal property. This type of study can be used to allocate tax dollars in multijurisdictional funding calculations or equalization by assuming that jurisdictions passing the performance review are assessing personal property at the general level of other classes of property analyzed with ratio studies. The study is completed by determining the amount of resources directed toward the assessment of personal property and reviewing appraisal and discovery methods. Therefore, performance reviews ascertain both the efficiency (cost and timeliness) and accuracy of assessment.

12.1.1 Personnel

Personnel should be assigned to personal property accounts. The number of accounts assigned to each personnel depends upon the complexity of the requirements, that is, inclusion of intangibles, inventories, household goods, agricultural products, motor vehicles, and complex exemptions. The appropriate number is also influenced by the amount of assistance provided by state or regional agencies.

12.1.2 Discovery

Identifying property escaping assessment, also known as discovery, is an important step in the performance review process. The jurisdiction must have the ability to discover the owners or users of taxable personal property within the jurisdiction. This is accomplished using numerous sources such as business licenses, online searches and by conducting a canvass. Similarly, property that is no longer part of the assessment roll must also be identified. The most comprehensive method of discovery of property is a well-designed property listing form completed by the property owner or their representative and should track both the additions and deletions of personal property for the period of assessment. A field canvass is also another comprehensive method of discovery that helps to identify property additions and deletions. (See Section 4 of the (Standard on valuation of personal property, 2018)).

12.1.3 Valuation

Personal property may be valued by using the three recognized approaches to value, namely, cost, income and sales comparison approaches to value depending upon the objective of the assessment. The most widely used approach in valuing personal property for ad valorem taxation is the cost approach. This requires using acceptable schedules and methods including properly submitted renditions, depreciation schedules published by nationally recognized valuation firms, market data from published valuation guides, and other generally accepted valuation methods.

Personal property should be valued at its current level of trade. Other crucial factors to consider are those that influence the value in use, including utility, usefulness to the owner, or the actual income produced (see Section 7 of (Standard on valuation of personal property, 2018)).

12.1.4 Verification and Auditing

Jurisdictions should establish statutes that contain enabling language for regulatory compliance and enforcement measures which include: (i) a requirement that property owners file personal property statements to the jurisdiction, (ii) giving assessors and their representative's authority to examine the property, books, papers, and accounts of taxpayer, and (iii) appropriate penalties for those who fail to file timely returns, file inaccurate information, or deny the assessor access to property and records.

Accuracy of personal property returns and reports should be verified by an audit program. Auditing is defined as a systematic method of validating the accuracy of the assessment by validating the data regarding property classifications, estimation of useful life, cost factor, percentage good, and methods used to ascertain replacement cost new. The audit program should provide coverage of the entire tax base regardless of the jurisdiction's reappraisal cycle. Evidence of the overall thoroughness of the jurisdiction may be indicated by the number of value change notifications and penalties imposed during the personal property assessment process (See Section 6 of the (International Association of Assessing Officers (IAAO), 2018)).

12.2. APPRAISAL RATIO STUDIES FOR PERSONAL PROPERTY

The appraisal ratio study is a numerical study that produces an estimate of the level of assessment of personal property by developing a ratio of assessment for property that is on the tax roll through the use of appraisals. The appraisal ratio study develops the relationship between validated market value of personal property (by audit and/or inspection) to the market value as established for the assessment roll for the same time period.

12.2.1 Assessment Ratio for Personal Property

Personal property market values are usually derived from appraisals using a replacement cost new (RCN) less depreciation.

A comparison of the depreciation schedules in use to nationally accepted schedules would enable the calculation of a ratio for property on the roll. A statistically sound process should be used to select a sample that is representative of personal property on the tax rolls. Such a sample can be based on the intended use of the ratio study in direct or indirect equalization.

12.2.2 Stratification

Proper stratification of personal property accounts should be done for greater statistical accuracy. Strata should be based on the type and value of personal property accounts. Stratification by type of account should occur first. Personal property accounts can be divided into residential (motor vehicles, boats, aircraft, and the like), agriculture, and business accounts. Further stratification can occur in residential and agricultural accounts but is necessary in business or commercial accounts. Business accounts are usually stratified by size into a minimum of four groups. Value ranges for these groups should be derived from the value ranges in the local market. One example would be small (less than $250,000), medium ($250,000 to $1 million), moderate ($1–$5 million), and large (greater than $5 million). Individual size of account by value can be determined based on the prior-year personal property tax roll.

12.2.3 Property Escaping Assessment

Personal property is particularly prone to escaping assessment. Some determination should be made about the portion of taxable personal property not on the assessment roll. However, estimates based on national averages are less meaningful at the local jurisdictional level. A policy should be established to uniformly correct escaped property from the assessment roll. The policy may consider such items as the value of the omission, the number of tax years to be revised, and the percentage of the omission to the value on the assessment roll for each years of the omitted property.

12.2.3.1. Identifying Personal Property Not on the Roll

Discovery tools can be used to determine accounts not on the roll for a sample area or group. Once the extent of the problem is identified, a projection can be made of the percentage of personal property not identified on the assessment roll.

The accepted method of determining the property omitted in taxpayer returns/reports is to audit the account (see Section 6 of the (Standard on valuation of personal property, 2018)). The audit results are applied back to the account value. The resulting fraction is property that is escaping taxation within that particular personal property account. If appropriate sampling techniques are used in selecting the accounts for audit, the resulting ratio is applied to the total roll to help determine the percentage of personal property escaping assessment within the jurisdiction.

12.2.3.2. Computing the Level of Valuation

The overall ratio is then determined by reducing the valuation ratio by the percent of property wholly or partially escaping taxation. For example, if the valuation level is found to be 90 percent and it is determined that 5 percent of personal property is escaping assessment, then the corrected level of assessment is the valuation level times the percentage of personal property assessed: $0.90 \times (1 - 0.05) = 0.855$. For indirect equalization, this calculation would result in a higher equalized value.

12.2.3.3. Ratio Study Performance Standards

Ratio study performance standards for personal property vary with local conditions. Hence, the jurisdictions have to comply with the performance standards established by the appropriate assessment administrator. Assessment administrators may be able to develop target standards based on an analysis of past performance or results in similar markets elsewhere. Such an analysis can be based on ratio study results for the past five years or more. Established target standards should reflect the nature of the properties involved, market conditions, and the availability of reliable market indicators.

13. CONSIDERATIONS FOR OVERSIGHT

In this standard, Oversight Agency refers to a completely independent oversight entity or separate internal division of a valuation organization responsible for quality assurance.
This section provides additional information specific to oversight and should be read in conjunction with the prior sections.

Oversight agencies perform ratio studies to monitor valuation performance and may establish independent and consistent parameters separate from those used by primary valuation offices. Common uses of these studies by oversight agencies may include:

- Advise and assist primary valuation offices.
- Equalize primary assessing agency valuations.
- Issue revaluation orders.

Principles

- There are differences between ratio studies used for oversight and those used by primary valuation offices.
- Oversight agency ratio study processes include stratification requirements for direct and indirect equalization.
- Various actions and roles of oversight agencies exist with regard to use of ratio studies.
- Independent valuations may be included to improve sample representativeness.
- Remedial actions should be based on pre-defined standards and ranges.

13.1. EQUALIZATION

Oversight agencies may use the results of ratio studies to equalize, directly or indirectly, valuations in taxing jurisdictions. Direct equalization is accomplished by an oversight agency which alters values determined by primary valuation agencies by ordering valuations within jurisdictions or property classes to be adjusted to market value or to the legally required level of valuation. Direct equalization can also involve adjusting valuations of centrally valued properties. When indirect equalization is used, valuations are not adjusted. Instead, indirect equalization involves an oversight agency estimating total taxable value, given the legally required level of assessment or market value. Indirect equalization allows proper distribution of intergovernmental transfer payments between different level governments despite different levels of valuation among jurisdictions or property classes. Equalization is not a substitute for revaluation.

When equalization is based on ratio study samples, sampling error must be taken into account. When confidence intervals include an acceptable range, equalization cannot be supported statistically. When confidence intervals fail to bracket official requirements, equalization actions are supported (see Section 7- Level of Valuation Statistics). Legal aspects of ratio studies, many of which relate to equalization, are discussed in Appendix G.

13.1.1 Direct Equalization

The advantage of direct equalization is that it can be applied to specified strata, such as property classes, geographic areas, and political subdivisions that fail to meet valuation level performance standards. Direct equalization also produces results that are generally more visible to the taxpayer and more clearly reduces perceived inequities between classes. For example, direct equalization allows proper and equal application of debt and tax rate limits and equitable partial exemptions.

Direct equalization involves use of adjustment factors, which produce effects mathematically identical to those derived through the application of "trending" or "index" factors, which are commonly used for value updating by primary valuation agencies. The most significant differences typically are the level of the jurisdiction originating the adjustments and the stratification of property to which the factors are applied. It is rare for equalization factors developed by oversight agencies to be applied to strata more specific than property class or broad geographic area. Often such factors are applied jurisdiction wide.

Although not a substitute for valuation or revaluation, direct equalization applied at the stratum level improves equality in effective tax rates between strata and lessens the effect of valuation practices that improperly favor one stratum over another. For example, assuming that all classes of property are to be valued at 100% of market value, without such equalization, in a case where residential property is valued at a median of 80% of market value, while commercial property is valued at a median of 100% of market value, residential property will pay a lesser percentage of its proper tax share than commercial property. Direct equalization mitigates this problem. However, such equalization cannot improve uniformity between properties within a given stratum. So, in the previous example, the median level for residential property can be adjusted from 80% to 100% of market value, but valuation disparities between individual residential properties will not be addressed. For this reason, revaluation orders should be considered as the primary corrective tool for uniformity problems, and direct equalization should be considered appropriate only if time or other constraints preclude such an approach.

13.1.2 Indirect Equalization

The most common use of indirect equalization is to enable proper intergovernmental funding distribution, particularly for school districts. Such equalization provides an estimation of the proper tax base (acknowledging statutory constraints such as agricultural use value) despite valuations that are higher or lower than legally required levels in certain jurisdictions. For example, if the valuation roll for residential property in a jurisdiction shows a value of $750 million, but a residential ratio study shows a valuation level of 75 percent, while the legally required level is 100 percent, an equalized value of $1,000 million could be computed ($750 million/0.75). This adjusted or equalized value would then be used to apportion payments or requisitions between higher and lower level governments.

Indirect equalization results in fairer funding apportionment because the overall valuation levels of the taxing jurisdictions tend to vary. If there were no equalization, the extent that a jurisdiction under- or overestimated its total tax base would result in over- or under-apportionment of funds. However, indirect equalization does not correct under- or overvaluation between classes of property within a jurisdiction. It usually adjusts only intergovernmental payments, is less visible to taxpayers, and often lacks checks and balances associated with direct equalization.

13.2. DESIGN OF STUDY FOR OVERSIGHT PURPOSES

The most important design consideration is that the study sample be sufficiently representative of the population of properties or the distribution of values in the jurisdiction under review. For direct equalization, the level of valuation for property classes or strata subject to such equalization is the primary area of interest, and the sample must be designed accordingly. Indirect equalization seeks to estimate the overall value of the population, so the sample must be representative of that overall value and must reflect the disproportionate influences of high value properties. For general performance monitoring, the study must be designed so that both level and uniformity statistics reflect those aspects of the underlying population.

13.2.1 Determining the Composition of Samples

In the design stage, the oversight agency must decide whether the ratio study sample should comprise sales, independent valuations, or a combination of the two. Each sample type has its advantages and disadvantages, as described below.

13.2.1.1. Sale Samples

The advantages of using sale samples include the following:

- Properly validated sales provide more objective indicators of market value than independent valuations;
- Using sales is much less expensive than producing independent valuations.

The disadvantages include the following:

- Difficulty in collecting sales data in jurisdictions without disclosure requirements;
- The oversight authority may not have control over the sales data collection and validation process;
- Influence of sales chasing can be difficult to detect or prevent;
- Samples of sales may not adequately represent the population of properties;
- An adequate sample size may not be achieved if sales data are scarce;

13.2.1.2. Independent Valuation Samples

Independent valuations can also be used instead of or combined with sales for ratio study samples. Advantages and Disadvantages of this approach are discussed in Appendix A.

13.2.2 Collection and Preparation of Market Data

The reliability of a ratio study depends in part on how accurately the sales or independent appraisals used in the study reflect market values. For sales-based studies, oversight agencies should conduct an independent sales verification and screening program if resources permit. Alternatively, oversight agencies should develop audit criteria to review data submitted to qualify sales, corroborate representativeness, and confirm adequate sample size. Oversight agencies that develop ratio studies from sales provided by local assessment jurisdictions should track the number of transfers obtained in different study periods. Quality control techniques can be used to measure market activity or to determine whether an assessor is fully reporting sales information.

Independent valuations used in ratio studies must reflect market values as of the assessment date being studied. To produce credible independent valuations, the oversight agency must be certain that property data used in developing its values are accurate.

13.2.3 Stratification for Equalization Studies

Predefined stratification is more transparent and enhances cooperation between the oversight agency and the jurisdiction valuing the property subject to equalization. In general, oversight agencies should not redefine the strata once they have been defined for equalization purposes, especially in the case of direct equalization. It is appropriate, however, to collapse strata to compensate for otherwise inadequate sample sizes, especially in the case of indirect equalization. If value stratification is necessary, predefined strata may not be practical. In addition, a revaluation or equalization order can be targeted for specific problem areas that cause noncompliance at a broader level of aggregation.

13.2.3.1. Stratification for Direct Equalization

Strata should be chosen consistent with equalization requirements. Statistical issues in the determination of strata include the size of the population and resulting strata, and the likely variability of the ratios in each stratum. Care must be taken not to over-stratify, that is, to create strata that are too small to achieve statistical reliability (see Section 5- Sample Selection and Timing of Studies and (Sherrill & Whorton, Jr, 1991)). No conclusion about stratum level or uniformity should be made from stratum samples that are unreliably small (resulting in unacceptably large margins of error) or where confidence intervals cannot be computed.

Ultimately, the degree of stratification is determined largely by available sales data, unless it is cost-effective and practical to add sufficient independent valuations. If sufficient sales or independent valuations are not available for a given stratum, it should be combined with similar strata.

Combination of strata with insufficient observations permits broader applicability of ratio study results and prevents ratio study analysis from becoming too focused on substrata with few sales or independent valuations. When jurisdiction or category wide equalization actions are required, reliability of component strata is not an issue.

13.2.3.2. Stratification for Indirect Equalization

Indirect equalization develops an estimate of full market value, but valuations of individual properties are not altered. Such studies can use a substantially different approach to stratification than ratio studies which are intended for performance evaluation or direct equalization. The purpose of stratification in this case is to minimize distortions due to different assessment levels, which can vary by property type, value range, geographic area, and other factors. A reasonable number of strata with small samples and larger margins of error can increase overall representativeness and may reduce the margin of error for the overall jurisdiction-wide sample.

The primary level of stratification should ordinarily be by major property type (e.g., residential, commercial, and vacant land). If circumstances permit, a secondary level of stratification also is recommended. When relying on the weighted mean, the secondary level of stratification (substrata) should normally be value range. Higher-value properties can sell with a different frequency than low value properties, and valuation levels can vary between high and low-value properties. As a result, high-value properties can be oversampled (or under sampled) and, because of their high value, can exert a disproportionate influence on the weighted mean and resulting estimated value. Value stratification reduces distortion of the weighted mean caused by over or under-representation of value strata with different levels of valuation. To properly develop and use value strata, the oversight agency needs each individual valuation in the study population. If detailed value information is not available, the oversight agency should work with primary valuation agencies to obtain sufficient information including the total value and number of properties in predetermined value categories.

In situations in which value stratification information is not available, or where property ratios are not significantly value-influenced, substrata can be created based on property subtype, geographic area, or other appropriate criteria. Stratification by these criteria corrects for differences in level of valuation between substrata. In large jurisdictions, sub-stratification by geographic areas generally is more appropriate for residential properties while sub-stratification by either geographic area or property subtypes (e.g., office, retail, and warehouse/industrial) can be appropriate for income-producing properties.

When relying on the median and when sample sizes permit, it is appropriate to stratify within property class by whichever property characteristic is most likely to capture differences in valuation levels. This characteristic can be geographic area, property subtype, or value range. Stratification by value range helps capture value-related differences in assessment levels which (unlike the weighted mean) are not reflected in the overall median.

13.3. TIMING AND SAMPLE SELECTION

Ratio studies made by oversight and equalization agencies should be conducted annually or to coincide with reassessment or equalization cycles. Where possible, ratio studies conducted by equalization agencies should use final values established at the primary valuation agency, inclusive of changes made by local appeal boards up to that time. However, if the primary valuation agency or appeal board "chases sales" or sets values in a manner that is dissimilar to the way other property values have been set, the sample may not be sufficiently representative and should not be used without careful investigation and necessary adjustment.

13.3.1 Date of Analysis and Related Considerations

When prior-year assessments are used to gauge current performance (to avoid sales chasing), the results should be adjusted for any revaluation activity or assessment changes that occurred in the population (net of new construction) between the prior and current years. Sale prices also should be adjusted to the assessment date to account for time trending. While this is an effective method for offsetting the effects of sales chasing on the level of valuation, caution should be used when drawing any conclusions about uniformity.

If the purpose of the study is equalization, using sales after the valuation date (adjusted for time as necessary) helps ensure the independence of assessed values and sales prices. A sales period spanning the valuation date can be used if measures are taken to ensure the independence of valuations determined after the earlier sales. This approach has the advantage of reducing the importance of time adjustments.

13.3.2 Required Sample Size

Because designing for sampling objectives and planning for resource allocation in ratio studies must occur well before final ratio data sets are available and ratio study statistics are calculated, decisions on critical input variables must be made well before their true values are known. For example, the sample size formulas ((Cochran, 1977); (Sherrill & Whorton, Jr, 1991); and (Gloudemans R. , 1999)) used to plan for specific margins of error and/or specific levels of confidence theoretically require, as input variables, the actual variation within the final ratio data sets (usually measured by the coefficient of variation). However, the actual variation in final ratio data sets is not known during the design and planning stage and, thus, the desired sample size must be projected based upon the best information available at the time of design and planning. This projection results in unavoidable forecast error and can result in the production of a higher or lower sample size than needed to reach sampling objectives. This issue is an accepted part of conducting ratio studies when it is necessary and important to attain a predetermined or uniform degree of precision. In other cases, it may be acceptable to use all available qualified sales. When predetermination of sample size is important, the variation in the ratio data set from the most recent time-period available can provide a reasonable estimate for the time-period under analysis.

13.3.3 Representativeness of Samples

The design and conduct of ratio studies require decisions that maximize representativeness within the constraints of available resources.

In many kinds of statistical studies, samples are selected randomly from the population and from within each stratum to maximize representativeness. Ratio study samples based on independently valued properties can be randomly selected. Because sales do not represent true random samples, care must be taken to maximize the representativeness of sales samples.

A ratio study sample is considered sufficiently representative for direct equalization and mass valuation performance evaluation when the distribution of ratios of properties in the sample reflects the distribution of ratios of properties in the population. A ratio study is considered sufficiently representative for indirect equalization when the distribution of property values in the samples reflects the distribution of property values in the population.
Sales from areas or substrata in which the number of sales is disproportionately large can distort ratio study results by weighting level and uniformity indicators toward whatever conditions exist in the overrepresented area. To alleviate this problem and create better representativeness, large samples can be further stratified by:

- randomly selecting sales to be removed;
- isolating the overrepresented groups into substrata;
- redefining the time-period for the overrepresented groups;
- weighting the data.

Most importantly, care must be taken to ensure that independent valuations reflect market value or value based on statutory constraints as of the appraisal date.

13.3.3.1. Maximizing Representativeness with Independently Valued Properties

When independently valued properties are added to sales used in ratio studies, the application of random sampling techniques can help ensure that valuation procedures used for the sampled properties are similar to the corresponding population. A well-designed random sampling plan also can help ensure that properties selected for independent valuation are not concentrated in areas of high sales activity or associated with property types with higher turnover rates in the market.

13.3.3.2. Extreme High-Value Properties

Assessment jurisdictions often contain unique, very high value properties (for example, properties that constitute more than 10 percent of the value of a property class) that cannot reasonably be combined with other properties for purposes of the ratio study. For indirect equalization, high-value parcels are especially important to maximize representativeness. For instance, consider a population consisting of 1,000 properties, 999 of which range in value from $20,000 to $750,000, and one that is valued at $1 billion (e.g., a power plant). If the intended use of the ratio study is to estimate the general level and uniformity of valuation in regard to the typical property, the stratified population of parcels need not include the $1 billion property. If the intended use of the ratio study is to estimate the total market value in the jurisdiction, however, exclusion of the power plant can distort the study.

Very high-value properties should not be ignored or assumed to be valued at the legal or general level for indirect equalization studies. An equalization agency can place very high-value property in a separate stratum to prevent distortion of the overall weighted mean or total estimated value. Provided there are sufficient numbers of properties in very high value strata, the oversight agency should perform a ratio study on these strata and use the results to adjust the value of the jurisdiction when warranted.

13.3.3.3. Outlier Ratios

Oversight agencies should consider the extent of sales verification when developing guidelines for trimming limits. In practice, this means that if an oversight agency derives sales data from assessing jurisdictions that may have already removed outliers from the sample, additional trimming may not be necessary (see Appendix C- Outlier Trimming Guidelines).

13.3.3.4. Value Outliers

When the weighted mean is used for indirect equalization or for equalization of centrally assessed property, a method that identifies high-value influential sales is recommended. Since an influential sale may not have an unusually low or high ratio relative to the rest of the sample, the definition of distortion is based on the principle that the point estimate calculated from the sample should not be statistically significantly different whether the suspect observation is in the sample or not.

To test for an influential sale, one approach is to remove it from the sample and compute the weighted mean and associated confidence interval. If the weighted mean of the sample lies outside the confidence interval calculated without the influential sale, then the sale is truly influential and is a candidate for further scrutiny, isolation in a separate stratum, or possible trimming. An additional approach is to identify the likelihood of occurrence of an extremely high assessed value in the probable distribution of all assessed values in the population. If the extremely high value represents for example, 10% of the total of all assessed value in the sample, it will carry a 10% weight with respect to the weighted mean which may overstate its influence on the population where it may only be 1% of the population (see (Dornfest & Chizewsky, 2017)).

These procedures test the presence of individual influential sales and are not intended to be used successively after deletion of a sale but can be applied to more than one apparent outlier at a time by leaving all other sales in the comparison group. Note, however, that the presence of multiple influential sales can indicate the start of a trend.

13.3.3.5. Outlier Trimming

Statistics calculated from trimmed distributions cannot be compared to those from untrimmed distributions or interpreted in the same way. This is especially problematic when making interjurisdictional comparisons. For this reason, oversight agencies may wish to promulgate uniform trimming procedures, based on sound statistical principles. Regardless of the chosen procedure, trimming of outliers must not occur more than once for any sample.

13.4. MEASURES OF VALUATION LEVEL FOR OVERSIGHT AND EQUALIZATION

The median is the generally preferred measure of central tendency for direct equalization, monitoring of valuation performance and evaluation of the need for a revaluation. The weighted mean is most appropriately used in indirect equalization when estimating the total value of the jurisdiction (see Table 10). When relying on the measure, outliers should be carefully reviewed (and deleted if appropriate), since they can strongly affect the weighted mean, particularly when they occur for high-value properties and in small samples. The mean should not be used for indirect equalization if there are measurable differences in appraisal level of high- and low-value properties. In data commonly containing outliers, the trimmed mean can be substituted for the mean (Gloudemans R. , 1999), chapter 3). See Appendix C for ratio outlier-trimming procedures.

Table 9 - Illustration of determining overall ratio for indirect equalization

Stratum	Total sample assessed value	Total sample sale price	Weighted mean	Total assessed value of stratum	Indicated market value of stratum
Residential	$ 3,000,000	$ 4,000,000	0.750	$ 600,000,000	$ 800,000,000
All other	$ 950,000	$ 1,000,000	0.950	$ 400,000,000	$ 421,000,000
Total				**$ 1,000,000,000**	**$ 1,221,000,000**

$$Overall\ ratio = \frac{\$\ 1{,}000{,}000{,}000}{\$\ 1{,}221{,}000{,}000} = 0.819$$

13.4.1 Overall Ratio for Combined Strata

The preferred approach for monitoring overall valuation performance and direct equalization is to weight the median ratio of each stratum based on the relative number of properties in the stratum. For indirect equalization, the weight assigned to a measure of central tendency of a stratum should be proportional to the share of that stratum's total estimated market value. Because the number of parcels bears only a loose relationship to total value, weighting by number of parcels is not appropriate for indirect equalization.

For indirect equalization, the preferred method of calculating the overall market value of a jurisdiction is as follows:

1. Divide the total value of each stratum by the stratum sample's measure of central tendency (see Section 13.4.2- Contrasting Measures of Valuation Level for Oversight and Equalization) to obtain an estimate of the total market value of taxable property in the stratum.
2. Sum the estimates of total stratum market value to obtain an estimate of the total market value of taxable property in the jurisdiction or class of property.
3. To obtain an overall weighted level of valuation, divide the total value of the jurisdiction or class of property by the estimated total market value (Table 9 contains a simplified example).

13.4.2 Contrasting Measures of Valuation Level for Oversight and Equalization

Table 10 summarizes the preferred measures of central tendency for the three broad purposes of indirect equalization, direct equalization, and the general monitoring of valuation performance.

Table 6- Preferred Estimators

	Indirect Equalization	Direct Equalization	Monitoring Performance
Median 1	-	X	X
Mean	-	-	-
Weighted Mean	X[1]	-	-
[1]: *Caution should be exercised when the sample contains value outliers or indicates value related bias, in which case the median should be substituted.*			

For indirect equalization, the preferred measure is the weighted mean (the measure used in Table 9). This helps achieve an accurate estimate of total value, the goal of indirect equalization. However, there are implicit difficulties in obtaining sales samples that are representative of all significant groups of properties with different ratios. The weighted mean can be disproportionately influenced by high-value properties, particularly in a small sales sample. This influence of high-value properties can be reduced through value stratification within the property class. Such value stratification helps capture value-related ratio differences, as well as improve representativeness, regardless of which measure of central tendency is used. If there are provable value-related ratio differences within strata, the weighted mean must be used since the median is incapable of capturing value-related differences. In cases in which value stratification is not practicable, equalization agencies may stratify by some proxy for value, such as neighborhood or property sub-class. If results appear distorted by non-representative high-value sales, outlier identification methods described in Appendix C should be employed.

While not conceptually preferred, the median can be used to prevent the disproportionate influence of high-value properties with outlier ratios. To be clear, although the median is not the conceptually appropriate measure, it nonetheless has the desirable property of smaller sampling variance and, in cases in which assessment regressivity/progressivity has not been found to be a significant concern, can provide an acceptable substitute for the weighted mean.

13.5. INDEPENDENT PROPERTY VALUATION-BASED RATIO STUDIES

Oversight agencies may conduct ratio studies by using independently determined value estimates . See Appendix A-Independent Value Estimate-based Ratio Studies.

13.6. RATIO STUDY STANDARDS FOR OVERSIGHT PURPOSES

Each oversight agency should have ratio study performance standards. These standards are suggested for jurisdictions in which current market value is the legal basis for valuation. In general, when state, provincial, or national standards are not met, revaluation or other corrective measures should be taken, or equalization procedures can be imposed. When an oversight agency orders such actions, the burden of proof should be on the agency to show that the standards have not been achieved.

All standards recommended in this section are predicated on the assumption that all practicable steps necessary to maximize representativeness and validity in the underlying ratio studies have been conducted.
When ratio studies are conducted for equalization purposes, confidence intervals and statistical tests can be used to determine whether it should be concluded at a given confidence level that appraisal performance or level requirements in a stratum (or jurisdiction) being tested meets or falls outside of mandated standards. Statistical tests can be used for comparisons among strata, provided the sample sizes are large enough that meaningful differences are not missed.

13.6.1 Level of Valuation – Oversight Use

Confidence intervals or statistical tests of significance should be used to determine whether the valuation level differs from the established goal in a particular instance.

A decision by an oversight agency to take some action (direct equalization, indirect equalization, revaluation) can have profound consequences for taxpayers, taxing jurisdictions, and other affected parties. This decision should not be made without a high degree of certainty that the action is warranted. Conversely, a decision not to take action when action is needed can have equally profound consequences. Oversight agencies should weigh all the options and consider the issues discussed below when developing or revising a level-of-valuation standard, and when developing equalization or other valuation oversight procedures.

13.6.1.1. Purpose of Level-of-Valuation Standard

Jurisdictions that follow the IAAO recommendation of annual revaluations should be able to develop mass valuation models that maintain an overall ratio level of 100 percent (or very near thereto). The local assessor may be required to observe revaluation cycles defined by a legal authority or public policy that can extend beyond one year. During extended cycles inflation or deflation can influence the overall ratio.

The purpose of a performance standard that allows reasonable variation from 100 percent of market value is to recognize uncontrollable sampling error and the limiting conditions that may constrain the degree of accuracy that is possible and cost-effective within a valuation jurisdiction. Further, the effect of performance standards on local assessors must be considered in light of expectations of public policy and resources available. For these reasons, oversight agencies may adopt performance standards for valuation level that allow some variance from the 100 percent goal of market value.

13.6.1.2. Recommended Valuation Level Standards for Direct and Indirect Equalization

The performance standard adopted by an oversight agency should be a range around the legally required level of valuation in a property class or an overall jurisdiction. This range should be 90 to 110 percent of the legally required level of valuation for direct equalization or revaluation, or 95 to 105 percent for indirect equalization. A smaller maximum range for indirect equalization is justified because taxpayers are not as comprehensively affected. Oversight agencies should adopt performance standards that are as close to the legally required level as can be justified given the local situation and taking into account the factors discussed herein.

In addition to the above valuation level standards, each class of property for which valuation level standards have been defined should not have measures of central tendency provably different by more than 10%. Both criteria must be met.

For example, if the valuation level for residential property is 0.93 and the valuation level for commercial property is 1.06, the jurisdiction is not in compliance with this requirement. This test should be applied only to strata subject to compliance testing. The oversight agency can conclude that this standard has been met if 90 percent (two-tailed) confidence intervals about the chosen measures of central tendency for each of the stratum fall within 10 percent of the overall level of valuation calculated for the jurisdiction or within 10% of the level of valuation of any strata. Using the above example, if the upper confidence limit for the level of residential property is 0.97 and the lower confidence limit for commercial property is 1.01, the two strata are within the acceptable range.

13.6.1.3. Confidence Intervals in Conjunction with Performance Standards

By themselves, the calculated measures of central tendency provide only an indication, not proof, of whether the valuation level meets the performance standard. So, the purpose of confidence intervals and similar statistical tests is to determine whether the valuation level differs from the established performance standard in a particular instance. A conclusion of noncompliance requires a high degree of confidence, thus a 90 percent (two-tailed) or 95 percent (one-tailed) confidence interval should be used, except for small or highly variable samples as described in Section 13.6.1.5- Adjustment for High Variability and Small Samples.

13.6.1.4. Decision Model

The oversight agency should determine whether the estimate is outside the acceptable range around the legal level of valuation with a specified degree of statistical significance. The chosen interval should overlap the performance standard range of 90 percent to 110 percent in the case of direct equalization or measuring valuation performance. For indirect equalization the chosen interval should overlap the performance standard range of 95 percent to 105 percent. If the confidence interval does not overlap any portion of the appropriate range, equalization is performed, or revaluation orders are issued. See Table 11 for an example of the direct equalization or valuation performance decision making process. When the interval includes a desired assessment level or a performance standard range around the desired level, equalization adjustments are not warranted. Similarly, when the interval includes a maximum allowable COD, reappraisal or other action to correct poor uniformity is not warranted.

Table 71- Ratio Study Standards and Decision Making-Direct Equalization or Valuation Performance Using Median 90%-110% Standard*

Case	Point Estimate	Confidence Interval (CI) Width (95%)	CI Overlaps Performance Standard Range	Point Estimate in Performance Standard Range	Equalization Action or Revaluation Order
1	92%	86% to 101%	Yes	Yes	No
2	88%	81% to 95%	Yes	No	No
3	84%	79% to 88%	No	No	Yes

**Example demonstrating application of standard at a 90% level of confidence*

13.6.1.5. Adjustments for High Variability and Small Samples

High variability, small sample size, or a combination of these factors often cause confidence intervals to become quite wide. Wide confidence intervals reflect the imprecision of the underlying statistic and can decrease the usefulness of performance measures. Also, wide confidence intervals can cause an inequitable situation in which jurisdictions with small samples and large variability are never subject to equalization or revaluation orders, while jurisdictions with larger samples and much less variability are more likely to be subject to such orders even though their valuation performance may be arguably better.

For these reasons, oversight agencies should consider expanding sample sizes by taking steps to increase the number of sales or by adding independently valued properties (see Appendix A). If the sample size cannot be increased, two options may be considered when the point estimate fails to achieve compliance, but the confidence interval overlaps the range of compliance:

- If a particular point estimate does not meet the standard for the current study cycle the oversight agency may reduce the level of confidence by 5% the following year. This may be followed by an annual stepwise reduction of 5%. Such a reduction may continue to a 70 percent level of confidence if the point estimate fails to meet the compliance threshold over this period of time. Corrective action would be imposed when a given year's confidence interval fails to include the performance standard range.
- The oversight agency may examine statistical point estimates over several study cycles. A jurisdiction that fails to meet a particular point standard for 5 consecutive years has a probability of less than 5% that compliance has been achieved, even if the confidence interval overlaps the compliance threshold every year. In such cases the oversight agency would impose corrective decisions based upon the point estimate.

13.6.1.6. Calculating Equalization Adjustments

If noncompliance with either direct or indirect equalization standards is indicated, the appropriate point estimate (statistic) measuring valuation level should be used to calculate adjustment factors, by dividing it into 100 percent.

13.6.2 Uniformity

Oversight agencies should establish uniformity standards for local assessment jurisdictions. Revaluation orders may be necessary to achieve these standards. Such orders should be dependent on statistically significant failures to meet standards. Suggested uniformity standards are presented in Table 7, and are defined in terms of CODs, but compliance should be evaluated using confidence intervals and other measures of statistical significance. The standards in Table 7 are intended to apply to ratio studies based on sales, not those based on independently valued properties, for which lower CODs typically are observed and expected.

In addition to horizontal uniformity standards based on the COD, oversight agencies should establish vertical equity standards (see Section 11.2.6 - Vertical Equity).

13.7. CAUTION ON USE OF RESULTS

Lack of independence between locally determined values and sale prices (sales chasing) or independent property valuation chasing can subvert attempts to improve equity (direct equalization) and result in incorrect distribution of funds between higher and lower-level governments (indirect equalization). To guard against these possibilities, oversight agencies should ensure that sold and unsold properties are valued similarly. Also, independent valuations used as substitutes for sales must reflect market value, and the oversight agency must take remedial measures in instances in which they do not (see Section 9- Considerations for Small Sample Situations and Appendix B- Sales Chasing Detection Techniques).

13.8. APPLICATION OF RESULTS

Once results are known, verified, and indicated with sufficient statistical significance, actions of oversight agencies as listed in the introduction to this section and further discussed below should be considered on the basis of these results. If assessments and oversight are the responsibility of a single entity, there should be provisions for independent review.

13.8.1 Advice and Assistance

Frequently the role of the oversight agency is to advise primary assessment jurisdictions of concerns about the quality of the assessments. Such advice should lead to follow-up to ensure that quality issues are resolved. However, the effectiveness of corrective measures cannot be ascertained until future ratio studies are done.

13.8.2 Equalize Local Valuations

Prior to issuing direct equalization orders oversight agencies should provide opportunities for primary assessment jurisdictions to take corrective actions or to comment on and review underlying data and ratio studies. In the case of indirect equalization affected jurisdictions should be given an opportunity to review the findings.

13.8.3 Revaluation Orders

Prior to issuing revaluation orders oversight agencies should communicate with local assessment jurisdictions to review quality concerns indicated by ratio studies. Grace periods should be considered by oversight agencies. Oversight agencies may consider significant or unusual market forces that may make a revaluation ineffective. Oversight agencies may consider the availability of reappraisal resources when issuing reappraisal orders.

APPENDICES

A. INDEPENDENT VALUE ESTIMATE-BASED RATIO STUDIES

Ratio studies may be conducted by using independent value estimates . A sampling plan can be designed to be more representative of the population in terms of property characteristics than a sales sample of the same size. Procedures for the analysis of personal property can be very similar, but are addressed separately (see Section 12).

The independent valuations should be made without knowledge of the assessor's value. Independent valuers should not be supplied with copies of the assessor 's valuation work sheets or model information. Supervisors should spot-check and review the work of staff valuers to ensure that the required independence is maintained. When the purpose of the ratio study is equalization or performance measurement, rather than internal quality assurance, the independent value estimates should not be revealed to the assessor until the assessor's values are final.

The use of independent value estimates, based on independent automated valuation models or single property appraisals, require adequately trained valuers and can be comparatively expensive. Many equalization or oversight agencies perform ratio studies in which sales and independent value estimates are combined. Executing a similar procedure can also benefit primary valuation offices. Furthermore, it can be possible to develop sales driven models for use in valuing a particular population of properties (excluding those not adequately represented in the underlying model) or randomly selected properties for ratio study purposes (Standard on automated valuation models, 2018).

A.1. Rationale

Independent value estimates can be used as indicators of market value. Independent valuations are performed by valuers who were not part of the group that has carried out the initial valuations. Independent value estimates include those derived from an independently created automated valuation model. Ratio studies based on such independently determined values are particularly useful for property classes with limited sales data, such as commercial and industrial real property and personal property (see (Property appraisal and assessment administration, 1990), Appendix 1-1 and (Gloudemans R. , 1999), chapter 60). In addition, independent value estimate-based ratio studies can be used for agricultural or other properties not valued on an ad valorem basis. In this case, the independent value estimates should reflect the use value or other statutory basis on which the properties are valued for property tax purposes.

A.2. Advantages and Disadvantages

Independent value estimate-based ratio studies have both benefits and limitations. Perhaps the greatest benefit of independent value estimate-based ratio studies is the ability to utilize effective sample selection practices that ensure a sufficiently sized and well-distributed sample. Such samples can, by design, accurately represent the entire property population, even in cases where there are insufficient sales, thus providing reliable and unbiased comparisons of assessed values to market values. If objectivity can be maintained, the independent value estimate-based ratio study avoids potential distortions due to systematic differences between valuations of sampled and unsampled properties. In addition, independent value estimates can be used to test for systematic differences between valuations of sold and unsold properties.

There are some important limitations to this approach. Beyond the expected added time and cost associated with conducting reviews and completing valuations to appropriate standards, there is the added challenge of completing these valuations where market data is scarce. This lack of direct market feedback requires additional scrutiny of the independent value estimates, both by those requesting the valuations and by the assessor, to ensure their accuracy and consistency.

A.3. Sample Selection and Resource Requirements

Sample selection and resource planning in independent value estimate-based ratio studies require knowledge of statistical sampling, estimation principles, and available resources. Judgment must be used because the determination of an adequate sample can require more information than is available during the design and planning phase, such as the actual variation within the final ratio data sets (see Section 9.1- Adequacy of a Given Sample

Size). Moreover, the cost of the study increases with the size of the sample. Therefore, the value of more reliable information must be balanced against the costs of obtaining that information.

In determining the size of the sample for each stratum, the following should be taken into consideration:

1. the required precision (typically measured by the margin of error) of the estimate of the valuation level, for example, ±0.05;
2. the required confidence level, for example, 95 percent;
3. the amount of dispersion expected in the final ratio data set;
4. the wastage associated with properties that cannot be efficiently valued or valuations that cannot be used for one reason or another (see (Gloudemans R. , 1999) chapter 6 for sample size formulas and required input variables; also see (Sherrill & Whorton, Jr, 1991)).

Once the desired size of an independent valuation sample has been determined, the individual properties that will constitute the sample should be selected using a statistically valid sampling plan. Stratified random sampling is preferred.

If value stratification is used, sample properties selected from value groups during resource planning can shift into other value groups before completion of the study, thus reducing the ultimate representativeness of the sample. Some intended valuation parcels may need to be removed from the sample when anomalous conditions are discovered such as environmental contamination (sufficiently reliable valuations may be prohibitively difficult or resource intensive) or when the independent valuer is not allowed access to the property. Any sample property that are voided or that shift from a stratum because of value changes should be replaced if possible.

Independent value estimate-based ratio studies (AMR based), as with sales ratio studies (ASR based), require informed, reasoned judgment to maximize sample representativeness and statistical reliability.

A.4. Data Requirements and Valuation Techniques

The valuation techniques selected for an independent value estimate-based ratio study should be consistent with accepted valuation principles and practices. The valuations should reflect the valuation date in question and should be well documented. Statistical software should be used as much as possible to expand analytical capabilities and perform calculations.

A.5. Independent Value Estimates from Automated Valuation Models

The valuations used in independent value estimate-based ratio studies can be based on CAMA and automated valuation model (AVM) techniques (Standard on automated valuation models, 2018). The models used must be developed independently from those used for assessment purposes. Adequate market and property characteristic data are required to develop reliable and defensible model estimates. If available, sales from a later period can be used to expand sample size. However, as in sales-based ratio studies, sales derived from primary valuation jurisdictions should be reviewed to ensure accuracy and validity. CAMA and AVM models have the advantage of reducing costs, permitting the use of larger, more representative samples. CAMA and AVM models developed for equalization must focus on the adequacy of overall, not individual, value estimates.

If the independent value estimates are used for ratio studies compared to the outcome of another automated valuation model it is advised that the model which creates the independent value estimates is based on a different statistical approach.

A.6. Independent Value Chasing

Independent value chasing can take two forms, either of which reduces or destroys the validity of the ratio study. The first occurs when an independent valuer knows the locally determined value and either consciously or unconsciously biases the independently determined value towards the primary assessing jurisdiction determined value. Independent valuers should not have access to the primary assessing jurisdiction's values or valuation work papers prior to completing their independent valuations. Also, independent value estimates should be reviewed and tested against the market.

The second form of independent value chasing occurs when the primary assessing jurisdiction knows which properties are in the ratio study sample and adjusts values on some or all of these properties to achieve better ratios without making similar adjustments to unsampled properties. This form of independent valuation chasing is similar to sales chasing and has similar consequences (see Appendix B- Sales Chasing Detection Techniques). Ratio study analysts should guard against this form of independent valuation chasing by withholding the release of sample information until the primary assessing agency values are final. If this form of independent valuation chasing occurs, the oversight agency can use primary assessing agency's values prior to adjustment to provide a more accurate representation of the population ratios.

A.7. Reviewing of Independent Value Estimates

Valuation supervisors should review valuation models or single-property valuations to ensure that standards are met. It also is good practice to include some recently sold properties in the sample being independently valued as a check on the validity of the methods being applied. In addition, the assessor must be afforded an opportunity to review the independent value estimates along with supporting documentation and to submit information supporting different value conclusions. If different value conclusions or factual information would materially affect the outcome of the study, a procedure for resolving conflicts, for example, by an independent review body, should be established.

A.8. Combining of Sales and Independent Value Estimates

Independently determined value estimates can be combined with valid sales in a ratio study. Using available sales adds objectivity to the study and reduces the required number of independent value estimates. On the other hand, combining sales and independent value estimates mixes two market indicators. If sales and independently determined values are combined, an analysis should be performed to test the consistency of measures of central tendency (mean and median) derived from the sales ratios compared to the same measures derived from the independent valuation-based ratios. Furthermore, the ratios should be distinguishable from one another in such a situation, careful labeling of the ratios must be done consistently (Hermans, Bidanset, Davis, & McCord, 2022). A Mann-Whitney test comparing values per unit or comparing ratios based on sales with those based on independent value estimates is appropriate for this purpose. Significant differences can result from several of the following conditions:

1. Sales have been chased;
2. Sales and independent valuations came from different geographic areas with different markets and different levels of valuation (maximize representativeness by stratifying);
3. Sales and independent valuations have different property characteristics that cause different levels of valuation;
4. All or some of the sales are invalid;
5. Outlier ratios are causing sale/ independent valuation ratio differences;
6. All or some of the independent valuations are inaccurate.

If none of the first five conditions listed above apply, the independent valuations should be tested against the market and revised as necessary (Wooten, 2003).

Variability measures (e,g, standard deviation, coefficient of dispersion, and interquartile range) computed on sales used in the sample may not always be similar to variability measures computed on independent valuations. Sales ratios reflect the vagaries of the marketplace. Ratios based on independent value estimates, on the other hand, come from comparing the results of one valuation model (e.g. the oversight agency's) to the results of another (e.g. the assessing office's). If both parties use mass valuation procedures, differences in valuations between the two models can be less than when compared with sales, particularly if the two models are too similar in their structure or are not truly independent of each other; in other words, variability measures based on independent valuation-based ratios can be lower than those based on sales ratios, even if the two groups represent properties with similar characteristics and similar degrees of valuation difficulty.

It is necessary that the use of independent value estimates is reflected on in the scope and design of the intended ratio studies. In presenting the results of a ratio study which includes the use of an independent value estimate it must be clear which ratio is used as underlying the calculation of the indicator.
For instance, ratios calculated between assessments and observed sale prices are referred to as the assessment to sales price ratio (ASR). Ratios calculated between assessment and the independent value estimate resulting from an automated valuation model are referred to as the assessment to model value ratio (AMR) (Hermans, Bidanset, Davis, & McCord, 2022). When a ratio study includes two or more differently constructed ratios, all ratios must be labelled in a way that each reported ratio is interpretable in a meaningful manner.

B. SALES CHASING DETECTION TECHNIQUES

Only when sold and unsold parcels are valued in the same manner and the data describing them coded consistently, can the statistics calculated in a sales ratio study be used to infer valuation performance for unsold parcels. If parcels that sell are selectively revalued or recoded based on their sale prices or some other criterion (such as listing price) and included in the ratio study, uniformity measures and inferences will likely be inaccurate (ratios, and thereby valuations will likely appear more uniform than they are). In this situation, measures of valuation level will be unsupportable and likely closer to market value than they are. Various forms of model overfitting can have the same practical effect of sales chasing.

Valuation officers and oversight agencies should consider implementing at least one of the detection procedures outlined in this Appendix. In addition, corrective procedures should be adopted when detection is indicated. In some cases, access to assessment information for all properties is necessary to perform the suggested techniques. Oversight agencies may not have such access and therefore may be more limited in terms of which detection techniques can be followed. Regardless, they may employ detection techniques, such as those described in sections B.3 and B.4, which do not require such comprehensive valuation information.

B.1. Comparison of Value Changes

If sold and unsold properties within a specified group are not valued in the same way, their valuations will likely result in different measures of central tendency of the percentage change from year to year. In determining the appropriate measure of central tendency, the median is preferred over the mean. However, there should also be a review of the proportion of sold and unsold properties with no value change. If this proportion is high in the unsold group, the median value change in that group may not be meaningful, in which case the mean should be used. On a deeper level, the distributions of the percent changes for sold and unsold properties can be compared.

In making this comparison, it is essential to remove properties that have undergone a substantive physical change or a change in use over the time period, as well as other properties that do not belong in such a comparison (e.g. agricultural use values that are not set by the valuation agency or values that changed due to an appeal / value stipulation). Such properties' value changes could bias this test towards detecting selective valuation, if included.

Statistical significance in the absence of practical significance may be moot. In large samples, small differences in the magnitude of value changes on sold and unsold parcels can be proven to be statistically significant, yet the actual differences may be slight. Therefore, it is prudent to establish some reasonable tolerance in addition to statistical significance. For this reason, it is recommended that statistically significant differences in valuation changes of sold and unsold properties greater than 5% likely indicate sales chasing that warrants remedial action (e.g., changes of 10 percent for sold properties and 4 percent for unsold properties). Such tolerance applies to other detection techniques discussed below. Useful tests of statistical significance include the Mann-Whitney U and T-test as indicated in B.2.

Value change comparisons can be visualized to pair with heuristic or statistical conclusions. Visualizations, such as boxplots, can compare value change percentages across sold and unsold property groups. These visualizations can be repeated across multiple data strata.

B.2. Comparison of Unit Values

If sold and unsold parcels are not valued equally, median or average unit values (for example, value per square foot) could be dissimilar. Appropriate tests (Mann Whitney U or *T*-test) can be conducted to determine whether differences are significant. The distributions of the unit values can also be compared statistically.

B.3. Split Sample Technique

In this technique, two ratio studies are performed, one using sales that occurred before the point at which sales would have been known by the valuation agency and one using sales that occurred after that point, both adjusted for date of sale as appropriate. A practical dividing point may be sales that occur after the last date at which the valuation agency can adjust values other than through appeal. As sales were not previously used, models and values can no longer be adjusted prior to conducting the ratio study. Taking into account appropriate market condition (i.e. time) adjustments, the results of the two studies should be similar. Sales chasing is indicated if a given statistic in the first study is statistically significantly better than in the second study and that the difference is practically meaningful. In such a case, the given statistic in the second study is still valid but the statistic in the first study should be rejected.

B.4. Comparison of Observed versus Expected Distribution of Ratios

Assuming the ratio studies are based on sales that have been properly adjusted for time and other factors, a strong indication of the likelihood of "sales chasing" can be obtained by computing the proportion of ratios that would be expected to fall within a particular narrow range of the mean given the lowest likely standard deviation (although this depends somewhat on the assumption of a normal distribution).

For example, with a standard deviation of 5 percent given a normal distribution, about 32 percent of the ratios would be expected to fall within ±2 percent of the mean (for example, between 98 and 102 percent, given a mean of 100 percent). Except in highly constrained or well-behaved real estate markets, many valuers consider such a low standard deviation, corresponding approximately to a COD of 4 percent, to be unachievable. Regardless of the distribution of the ratios, the likelihood is extremely low that there would be a sufficiently representative sample with more than this proportion of ratios in such a narrow range. If such is the case, "sales chasing" is a likely conclusion.

Although samples may not be normally distributed, in which case equivalently precise proportions of expected ratios around the median cannot be determined, the 32 percent concentration is very conservative. Finding such a high concentration of ratios around any measure of central tendency is a strong indicator of sales chasing or, at least, of a non-representative ratio study. In addition, when the distribution of ratios is bimodal or multimodal, similar significant concentrations of ratios around these modes can indicate selective revaluation.

Table 13 demonstrates the conservative nature of the 32 percent concentration. If the minimum achievable COD is, in fact, higher than 4 percent for the strata or property class being analyzed, then even lower concentrations could indicate sales chasing, and previously discussed investigative procedures should be instituted. One disadvantage to this procedure is that it can be misleading when applied to small samples. Therefore, the method should not be employed for sample sizes less than 30.

Even when critical proportions of ratios shown in Table 13 are exceeded, further investigation should be conducted before concluding that sales chasing has occurred.

B.5. Mass Appraisal Techniques

Provided sales are sufficient in number, valuation and oversight agencies can develop mass valuation models to apply to a random sample of unsold properties or to the population of properties that are represented by the sold properties. An independent multiple regression or other automated calibration techniques can be used to develop the models. A ratio study is then conducted for the unsold parcels by substituting sale price, the denominator of the sale ratio, with the values predicted by the independent models as they are used as indicators of market value for the unsold parcels. This approach has the following advantages:

- It is objective and rooted in the market.
- The models can be reviewed for sufficient reliability before being applied to the unsold parcels.
- The technique yields measures of central tendency, which can be compared against those produced by the sales ratio study and tested for compliance with standards for the level of valuation.
- The technique takes the form of an independent valuation ratio study but avoids the time and expense of single -property appraisals.

Reliability of this method depends on the accuracy and independence of the mass valuation models used to generate the value estimates. The models must be consistent with valuation theory and reviewed for sufficient reliability by examining goodness-of-fit statistics. The models should be independent of those used for valuation purposes.

To effectively detect sample-induced bias—such as that caused by sales chasing—an independent model should be developed that differs from the primary model in key respects. The second model should use a different modeling structure (e.g., linear regression versus gradient boosting) to ensure sensitivity to different forms of distortion. It should ideally be constructed by a separate analyst or team to avoid unintentional replication of assumptions. The underlying data should be handled independently, either through random sample splits or by structuring the second dataset to include or treat sales differently (e.g., blinding sales status), to break any feedback loop between modeling and sample selection. Variable selection should also differ, emphasizing (for example) model stability and generalizability over precision. These safeguards help ensure that the second model acts as a meaningful diagnostic tool, capable of revealing sample-induced bias that may remain hidden when only a single model structure and dataset are used. Structural differences also increase the likelihood of detecting model-induced bias, which may compound sample issues and further obscure equity concerns.

B.6. Independent Value Chasing Detection

As outlined in Appendix A, statistical inferences about the level and uniformity of the population of properties under review can also be distorted when ratio studies are based in part on proxy values determined independently by oversight agencies. In addition to other detection methods indicated in this appendix, statistically significant differences between ratios of independently valued properties and selling properties in the same strata should be analyzed. If differences that also are of practical significance are detected, procedures found in Appendix A should be used to correct ratio study results or employed preemptively to preclude value chasing when independent values are developed.

Table 8- Example of critical ratio concentrations indicative of sales chasing or similar practices

Minimum achievable COD	Standard deviation assuming normal distribution and mean ratio	Critical proportion of ratios*	zscore based on ± 2% range (Absolute value)	Expected proportion of ratios below 0.98	Expected proportion of ratios below 1.02	Expected proportion between 0.98 and 1.02 (within ± 2% of central tendency)
1.6%	2.00%	69	1.0000	0.1587	0.8413	0.6826
4.0%	5.00%	32	0.4000	0.3446	0.6554	0.3108
5.0%	6.25%	26	0.3200	0.3745	0.6255	0.2510
6.0%	7.50%	22	0.2667	0.3949	0.6051	02102
7.0%	8.75%	19	0.2286	0.4110	05896	0.1801
8.0%	10.00%	16	0.2000	0.4207	0.5793	0.1586
10.0%	12.50%	13	0.1600	0.4364	05636	0.1272
12.0%	15.00%	11	0.1333	0.4467	05530	0.1063
14.0%	17.50%	10	0.1143	0.4545	05455	0.0910
16.0%	20.00%	8	0.1000	0.4602	0.5398	0.0796

* Given the assumption that the COD shown represents the minimum achievable COD for the property type, class, or strata being analyzed with the ratio study, sales chasing (or a similar distortive procedure) is very likely if the concentration of ratios with ± 2% of a measure of central tendency, such as the median or a mode, or 700%, equals or exceeds this value. This proportion is based on values of the standard normal distribution function and

assumption that sample size is greater than 30. The critical number equals the integer immediately exceeding the expected proportion.

C. OUTLIER TRIMMING GUIDELINES

C.1. Identification of Ratio Outliers

Outlier ratios are ratios that lay beyond the general distribution of the data. Although exactly where outliers begin is often unclear, obvious outliers can be identified for further investigation by a simple array of the ratios as well as by various graphs, e.g., histograms, boxplots, and scatter graphs of ratios plotted against sale date, size, and other property characteristics.

As discussed below, outliers occur for various reasons, including data errors and sales screening lapses that should be corrected if possible. Removing outliers without proper cause can mask legitimate, sometimes serious data or valuation issues. Thus, the principle in screening outliers for ratio analysis should be, first, to correct those found to be due to legitimate errors and, second, to remove (e.g. trim) only outliers that would compromise the representativeness of resulting ratio statistics. Outliers should never be removed simply to generate better ratio statistics, which serves only to destroy the representativeness and validity of the study. As further detailed below, after legitimate data or valuation model corrections are made, assessment agencies should have a policy of removing no more than a relatively small, fixed percentage of ratios as outliers, thresholds for which can depend on sample size, property type, and the adequacy of sales data and screening.

Statistical procedures can aid in the identification and handling of outliers. One recommended technique applicable to ratio studies, illustrated in Table 14, is based on the interquartile range (IQR) and has been shown to be robust for a wide variety of distributions (Iglewicz & Hoaglin, 1993); (Barnett & Lewis, 1994). The procedure identified ratios that fall outside given multipliers of the IQR. The two most common multipliers are 1.5, which can often be expected to flag approximately 5% of ratios, and 3.0, which can often be expected to flag approximately 1% of ratios which are the most extreme outliers. Other outlier identification procedures are found in statistical literature and can also be used. In any case, outlier identification and trimming should follow the sales validation process and precede the calculation of ratio statistics and related tests or analyses.

The example in Table 14 demonstrates the use of the 1.5 X IQR procedure to identify outlier ratios. The distribution of ratios often is skewed to the right; therefore, it is advisable to apply an appropriate transformation to the ratios to make their distribution more symmetric prior to applying the IQR method. For example, the use of logarithmic transformations tends to achieve greater symmetry and identify fewer high and more low ratios as outliers. IQR-based outlier identification usually works well when sample sizes are sufficient to preclude the aberrant results that can be expected when this procedure is applied to very small, highly variable samples.

C.2. Scrutiny of Identified Outliers

The preferred method of handling an outlier ratio is to subject it to additional scrutiny to determine whether the sale is a non-market transaction or contains an error in fact. If an error can be corrected (for example, data entry), then it should remain in the sample so long as it is no longer a ratio outlier. If the error cannot be corrected or if inclusion of the identified outlier would reduce sample representativeness, the sale should be excluded.

C.3. Analytical Use of Identified Outliers

After identification, scrutiny, and correction of errors associated with outliers, the procedure can be run again to identify any remaining apparent outliers. If outlier ratios tend to be concentrated in certain areas or other subsets of the sample, they can point directly to systematic errors in the appraisal process and should be flagged for subsequent analysis.

C.4. Outlier Trimming

Once outliers have been identified and scrutinized and any errors resolved, the next step is to exclude (i.e. trim) those that may unduly influence calculated statistical measures. Recognized statistical procedures should be used, and the number of trimmed ratios should not exceed limits found in C.5. For cautionary notes on trimming small samples, see (Tomberlin, 2001) and (Hoaglin, Mosteller, & Tukey, 1983). An example of such trimming is found in Table 15. However, trimming outliers using arbitrary limits, for example, eliminating all ratios less than 50 percent or greater than 150 percent, tends to distort results and should not be employed. Outlier trimming should only occur as the result of a single procedure, not as the cumulative result of repetitive rounds.If outliers are to be considered for removal, the analyst can select a procedure to trim all or just the extreme or influential outliers (see Table 15). If a trimming method has been used to reject ratios from the sample, this fact must be stated in ratio study reporting, and outliers removed should be identified. Outlier trimming is not mandatory; however, if outlier-trimming procedures are not used, sales with extreme or influential ratios must be thoroughly validated and determined to be highly trustworthy observations because they can play a pivotal role in the ratio study outcome.

Using the data in Table 14, the following procedure identifies outlier ratios that fall more than 1.5 times beyond the range of the middle 50 percent of the arrayed sample.

Table 9- A Distribution-Free Method for Locating Outliers- Data before trimming

Rank	Ratio (A/S)
1	0.611
2	0.756
3	0.762
4	0.853
5	0.867
6	0.909
7	0.925
8	0.944
9	1.014
10	1.052
11	1.178
12	1.367
13	1.850
14	2.500
Median Ratio	**0.935**
COD	**32.271**

Steps to locate trim boundaries:

1. *Locate the first quartile point formula to locate the first quartile*:
   (0.25 x number of ratios) + 0.25
   (0.25x14 ratios) + 0.25 = 3.75
   3.75 is three-quarters between the third and fourth ranked ratios.
   Ratio 3 = 0.762
   Ratio 4 = 0.853
   Three-quarters between = (0.853 -0.762) x 0.75 = 0.068
   The first quartile point = 0.762 + 0.068 = 0.830
2. *Locate the third quartile point*
   Formula to locate the third quartile
   (0.75 x number of ratios) + 0.75
   (0.75x14 ratios) + 0.75 = 11.25

11.25 is one-quarter between the eleventh and twelfth ranked ratios.
Ratio 11 = 1.178
Ratio 12 = 1.367
One-quarter between = (1.367 -1.178) x 0.25 = 0.047
The third quartile point = 1.178 + 0.047 = 1.225

3. *Compute the interquartile range*
The distance between the first and third quartile = interquartile range
1.225 -0.830 = 0.395
4. *Establish the lower boundary*
Lower trim point = first quartile -(interquartile range x 1.5 or 3.0)
0.830 -(0.395 x 1.5) = 0.238
5. *Establish the upper boundary*
Upper trim point = (interquartile range x 1.5 or 3.0) + third quartile
(0.395 x 1.5) + 1.225 = 1.818

Outliers identified: 1.850, 2.500

Table 10- Effects of Outlier Trimming- Data after 1.5x trimming

Rank	Ratio (A/S)
1	0.611
2	0.756
3	0.762
4	0.853
5	0.867
6	0.909
7	0.925
8	0.944
9	1.014
10	1.052
11	1.178
12	1.367
Median Ratio	**0.917**
COD	**15.649**

C.5. Trimming Limitations

In the process of outlier trimming, the goal should be to remove observations responsibly. Outlier trimming can be an important tool in not allowing atypical observations to have an undue influence on calculated statistics. Trimming too many observations, however, may make the same calculated statistics seem better than they really are and may reduce sample representativeness. To that end, it is appropriate to set maximum limits for outlier trimming. For samples of less than 30, no more than 10 percent (20 percent in the most extreme cases) of the ratios should be removed. For larger samples, this threshold can be lowered to 5 to 10 percent depending on the distribution of the ratios and the degree to which sales have been screened or validated. Trim limits should generally be developed in consideration of the extent of sales verification, with a greater percentage trimmed when a thorough and properly conducted sales verification process has not taken place.

Also, some samples have low CODs and tight ratio frequency distributions, possibly due to unusually homogeneous populations of property. In such cases the IQR could identify a large proportion of the sample as “outliers” even though the difference between these sales in ratio terms is very small. If blindly followed, this can unnecessarily reduce sample size, eliminating representative sales and lowering statistical significance of results. For this reason, sales flagged by the IQR method should not be automatically excluded in such a low dispersion situation, but instead be subject to the maximum trimming limits outlined in this section.

C.6. Reporting Trimmed Outliers and Results

Ratio study reports or accompanying documentation should clearly state the basis for excluding outlier ratios. Statistics calculated from trimmed distributions, obviously, cannot be compared to those from untrimmed distributions or interpreted in the same way.

C.7. Handing of Value Outliers

While this section focuses on ratio outliers, values that make up the ratio can also be outliers that may have a distorting effect on certain statistics like the weighted mean. Guidance from this section such as outlier detection methods, scrutiny, trimming, and trimming limitations can be applied to value outliers as well. However, in order to prevent the inappropriate or unbalanced removal of value outliers, data should be transformed before the application of the IQR approach or an alternative method should be used that is based on ranking sales by assessed value or sale price. Alternatively, a more representative weighted mean (such as for indirect equalization) can be developed by identifying and trimming outliers found using specific value outlier identification procedures presented in Section 13.3.3.

D. CONFIDENCE INTERVAL CALCULATIONS FOR MEASURES OF CENTRAL TENDENCY AND THE COEFFICIENT OF DISPERSION

A confidence interval (CI) is a range of values that contains the true population parameter (such as a mean, median or proportion) with a specified level of confidence. It provides an estimate along with a margin of error to account for variability in the data.

For example, a **90% confidence interval** means that if we repeated the sampling process many times, 90% of the calculated intervals contain the true population value. A wider interval suggests more uncertainty, while a narrower interval indicates greater precision in the estimate. The confidence level (e.g., 90%, 95%, or 99%) reflects how sure we can be that the true value falls within the range.

This appendix outlines the steps to calculate confidence intervals for measures of central tendency as well as the Coefficient of Dispersion (COD). Two separate techniques are described to determine COD compliance with COD standards. Alternatively, compliance with COD standards may also be determined by estimating confidence intervals using bootstrapping methods.

For more information on calculating measures of central tendency and the COD, refer to Section 7 and Section 8, respectively.

D.1 Mean Confidence Interval

The formula for the confidence interval about the mean assumes a normal distribution. It is calculated with the following steps:

1. Calculate the standard deviation of the sample(s)
2. Determine the sample size (n) and calculate the degrees of freedom (n-1)
3. Determine the applicable confidence level
4. Identify the critical value of t for a two-tailed test using a Critical Value Table (t)
5. Multiply the critical value by the standard deviation
6. Divide by the square root of the sample size (n)
7. Determine the upper limit of the confidence interval by adding the result to the mean ratio
8. Determine the lower limit of the confidence interval by subtracting the result from the mean ratio

Mathematically, the general formula for the mean confidence interval is as follows:

$$\bar{x} \pm t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$$

where $\bar{x}$ represents the mean (e.g., the mean ratio in a ratio study)

For a 95% confidence interval, set t = 1.96. For a 90% confidence interval, set t = 1.645. These t constants assume a sample size (n) of greater than 30. For smaller sample sizes, there is more uncertainty and wider confidence intervals. t-values to be substituted can be found in standard “Student’s” t tables or in Table 19 in Appendix D.5.

D.2 Median Confidence Interval

Unlike the mean, the median confidence interval does not depend on the assumption of a normal distribution, and it is not nearly as affected by outlier ratios as noted in Section 7. When the sales sample is greater than 30 sales, the median confidence interval is calculated using the following steps

1. Array the ratios in ascending order and rank accordingly
2. Identify the ranks of ratios, j, corresponding to the upper and lower confidence limits for a 95% confidence interval as follows. For a 90% confidence interval, replace 1.96 with 1.645.
   a. If the sample size (n) is odd:
   $$j = \frac{1.96 \times \sqrt{n}}{2}$$
   where $\sqrt{n}$ is the square root of the sample size n.

   b. If the sample size (n) is even:
   $$j = \frac{1.96 \times \sqrt{n}}{2} + 0.5$$
   where $\sqrt{n}$ is the square root of the sample size n.

3. If the result from step 2 is not an integer, round up to the next integer.
4. Return to the array and locate the median.
5. Using the result from step 3, count up and down the array from the median to find the ratios which correspond as the upper and lower confidence limits for the desired confidence interval.

See D.4 for steps to calculate median confidence intervals for small sample sizes, less than or equal to 30 observations.

D.3 Weighted Mean Confidence Interval

The Confidence Intervals for the weighted mean are calculated as follows:

1. Calculate the weighted mean ratio per Section 7.3.
2. For each ratio, calculate the weighted variance by subtracting the weighted mean ratio from step 1, and then square the result. Then multiply each result by the weight for that ratio.
3. Sum the weighted variances from step 2.
4. Calculate the weighted standard deviation by dividing the total weighted variance from step 3 by the total of the weights minus the sum of the squared weights divided by the total of the weights. Then take the square root of this value.
5. Determine the effective sample size by calculating the sum of the weights and dividing by the sum of the squared weights.
6. Calculate the degrees of freedom by subtracting one from the effective sample size calculated in step 5.
7. Select the desired confidence level, such as 90% or 95%.
8. Find the critical value of t for a two-tailed test using a t-distribution table and the degrees of freedom from step 6.
9. Multiply the critical t-value by the weighted standard deviation.
10. Divide the result by the square foot of the effective sample size to calculate the margin of error.
11. Add the margin of error to the weighted mean from step 1 to determine the upper limit of the confidence interval.

12. Subtract the margin of error from the weighted mean from step 1 to determine the lower limit of the confidence interval.

See Gloudemans & Almy 2011 p. 267 Table 10.2 for a sample calculation of a 95% confidence interval for the weighted mean.

D.4 Median Confidence Intervals for Small Sample Sizes

For small samples, a formula based on the binomial distribution is required to calculate a confidence interval around the sample median. This method should be used for small samples with up to 30 observations. For larger samples, the method as outlined in section D.2 of this appendix may be used.

A **binomial distribution** describes the probability of getting a certain number of "successes" in a fixed number of independent trials, where each trial has only two possible outcomes (success or failure) and the probability of success stays the same. For example, if you flip a fair coin 10 times, it can help determine the likelihood of getting exactly 6 heads. In real estate, the binomial distribution can be used in a ratio analysis to estimate the probability of a certain number of properties having an assessed value within an acceptable range of their sale price, assuming each property has the same chance of meeting the criteria.

Tables 16 and 17 provide the parameters for a formula based upon the binomial distribution to develop the lower and upper median confidence interval estimates (Clapp, 1989). R^i is the ratio in an array ranked from the lowest to the highest (sorted in ascending order). Each row represents the sample size with the corresponding formula to be utilized.

Each confidence interval boundary typically falls between two ratios in the array. The interpolation factor is multiplied by the ratio value and the two are added together to obtain a weighted average.

Table 16 – 90% Confidence Intervals

n	Lower Bound	Upper Bound
5	$.8800 \times R^1 + .1200 \times R^2$	$.8800 \times R^5 + .1200 \times R^4$
6	$.6333 \times R^1 + .3667 \times R^2$	$.6333 \times R^6 + .3667 \times R^5$
7	$.2286 \times R^1 + .7714 \times R^2$	$.2286 \times R^7 + .7714 \times R^6$
8	$.8643 \times R^2 + .1357 \times R^3$	$.8643 \times R^7 + .1357 \times R^6$
9	$.5667 \times R^2 + .4333 \times R^3$	$.5667 \times R^8 + .4333 \times R^7$
10	$.1067 \times R^2 + .8933 \times R^3$	$.1067 \times R^9 + .8933 \times R^8$
11	$.7855 \times R^3 + .2145 \times R^4$	$.7855 \times R^9 + .2145 \times R^8$
12	$.4282 \times R^3 + .5718 \times R^4$	$.4282 \times R^{10} + .5718 \times R^9$
13	$.9558 \times R^4 + .0442 \times R^5$	$.9558 \times R^{10} + .0442 \times R^9$
14	$.6511 \times R^4 + .3489 \times R^5$	$.6511 \times R^{11} + .3489 \times R^{10}$
15	$.2217 \times R^4 + .7783 \times R^5$	$.2217 \times R^{12} + .7783 \times R^{11}$
16	$.8261 \times R^5 + .1739 \times R^6$	$.8261 \times R^{12} + .1739 \times R^{11}$
17	$.4603 \times R^5 + .5397 \times R^6$	$.4603 \times R^{13} + .5397 \times R^{12}$
18	$.9735 \times R^6 + .0265 \times R^7$	$.9735 \times R^{13} + .0265 \times R^{12}$
19	$.6480 \times R^6 + .3520 \times R^7$	$.6480 \times R^{14} + .3520 \times R^{13}$
20	$.2072 \times R^6 + .7928 \times R^7$	$.2072 \times R^{15} + .7928 \times R^{14}$
21	$.8048 \times R^7 + .1952 \times R^8$	$.8048 \times R^{15} + .1952 \times R^{14}$
22	$.4156 \times R^7 + .5844 \times R^8$	$.4156 \times R^{16} + .5844 \times R^{15}$
23	$.9413 \times R^8 + .0587 \times R^9$	$.9413 \times R^{16} + .0587 \times R^{15}$
24	$.5884 \times R^8 + .4116 \times R^9$	$.5884 \times R^{17} + .4116 \times R^{16}$
25	$.1203 \times R^8 + .8797 \times R^9$	$.1203 \times R^{18} + .8797 \times R^{17}$
26	$.7371 \times R^9 + .2629 \times R^{10}$	$.7371 \times R^{18} + .2629 \times R^{17}$
27	$.3161 \times R^9 + .6839 \times R^{10}$	$.3161 \times R^{19} + .6839 \times R^{18}$

n	Lower Bound	Upper Bound
28	.8687 x R^{10} + .1313 x R^{11}	.8687 x R^{19} + .1313 x R^{18}
29	.4831 x R^{10} + .5169 x R^{11}	.4831 x R^{20} + .5169 x R^{19}
30	.9876 x R^{11} + .0124 x R^{12}	.9876 x R^{20} + .0124 x R^{19}

Table 17 – 95% Confidence Intervals

n	Lower Bound	Upper Bound
6	.9000 x R^{1} + .1000 x R^{2}	.9000 x R^{6} + .1000 x R^{5}
7	.6857 x R^{1} + .3143 x R^{2}	.6857 x R^{7} + .3143 x R^{6}
8	.3250 x R^{1} + .6750 x R^{2}	.3250 x R^{8} + .6750 x R^{7}
9	.9222 x R^{2} + .0778 x R^{3}	.9222 x R^{8} + .0778 x R^{7}
10	.6756 x R^{2} + .3244 x R^{3}	.6756 x R^{9} + .3244 x R^{8}
11	.2873 x R^{2} + .7127 x R^{3}	.2873 x R^{10} + .7127 x R^{9}
12	.8936 x R^{3} + .1064 x R^{4}	.8936 x R^{10} + .1064 x R^{9}
13	.6056 x R^{3} + .3944 x R^{4}	.6056 x R^{11} + .3944 x R^{10}
14	.1659 x R^{3} + .8341 x R^{4}	.1659 x R^{12} + .8341 x R^{11}
15	.8218 x R^{4} + .1782 x R^{5}	.8218 x R^{12} + .1782 x R^{11}
16	.4827 x R^{4} + .5173 x R^{5}	.4827 x R^{13} + .5173 x R^{12}
17	.9899 x R^{5} + .0101 x R^{6}	.9899 x R^{13} + .0101 x R^{12}
18	.7076 x R^{5} + .2924 x R^{6}	.7076 x R^{14} + .2924 x R^{13}
19	.3059 x R^{5} + .6941 x R^{6}	.3059 x R^{15} + .6941 x R^{14}
20	.8835 x R^{6} + .1165 x R^{7}	.8835 x R^{15} + .1165 x R^{14}
21	.5479 x R^{6} + .4521 x R^{7}	.5479 x R^{16} + .4521 x R^{15}
22	.0697 x R^{6} + .9303 x R^{7}	.0697 x R^{17} + .9303 x R^{16}
23	.7381 x R^{7} + .2619 x R^{8}	.7381 x R^{17} + .2619 x R^{16}
24	.3373 x R^{7} + .6627 x R^{8}	.3373 x R^{18} + .6627 x R^{17}
25	.8958 x R^{8} + .1042 x R^{9}	.8958 x R^{18} + .1042 x R^{17}
26	.5481 x R^{8} + .4519 x R^{9}	.5481 x R^{19} + .4519 x R^{18}
27	.0677 x R^{8} + .9323 x R^{9}	.0677 x R^{20} + .9323 x R^{19}
28	.7221 x R^{9} + .2729 x R^{10}	.7221 x R^{20} + .2729 x R^{19}
29	.3063 x R^{9} + .6937 x R^{10}	.3062 x R^{21} + .6937 x R^{20}
30	.8709 x R^{10} + .1291 x R^{11}	.8709 x R^{21} + .1291 x R^{20}

Example

Table 11 - Data for Small Sample Median Confidence Interval Sorted by Ratio

Rank	Parcel #	Appraised Value	Sale Price*	Ratio
1	9	$87,200	$138,720	0.629
2	10	38,240	59,700	0.641
3	11	96,320	146,400	0.658
4	12	68,610	99,000	0.693
5	13	32,960	47,400	0.695
6	14	50,560	70,500	0.717
7	15	61,360	78,000	0.787
8	16	47,360	60,000	0.789
9	17	56,580	69,000	0.820
10	18	47,040	55,500	0.848
11	19	136,000	154,500	0.880

Rank	Parcel #	Appraised Value	Sale Price*	Ratio
12	20	98,000	109,500	0.895
13	21	56,000	60,000	0.933
14	22	159,100	168,000	0.947
15	23	128,000	124,500	1.028
16	24	132,000	127,500	1.035
17	25	160,000	150,000	1.067

**Or Time Adjusted Sale Price*

Using data from Table 16 (n = 17 ratios) and the parameters from Table 17, the 95% confidence interval for the median is calculated as follows:

Lower bound:

Step 1:	Identify sample size:	n = 17							
Step 2:	Go to 95% confidence interval **Lower Bound** table where n = 17								
	17	0.9899	x	R^5	+	0.0101	x	R^6	
Step 3:	Select Ratio 5	= .695							
	Select Ratio 6	= .717							
Step 4:	Place Ratios 5 & 6 in formula								
	17	0.9899	x	0.695	+	0.0101	x	0.717	
Step 5:	Calculate Lower Bound:								
		0.6880	+	0.0072	=	**0.6952**			
Answer:	**0.6952**								

Upper bound:

Step 1:	Identify sample size:	n = 17							
Step 2:	Go to 95% confidence interval **Upper Bound** table where n = 17								
	17	0.9899	x	R^{13}	+	0.0101	x	R^{12}	
Step 3:	Select Ratio 13	= .933							
	Select Ratio 12	= .895							
Step 4:	Place Ratios 13 & 12 in formula								
	17	0.9899	x	0.933	+	0.0101	x	0.895	
Step 5:	Calculate Lower Bound:								
		0.9236	+	0.0090	=	**0.9326**			
Answer:	**0.9326**								
95% Median CI	**(0.6952, 0.9326)**								

As noted in *Property Appraisal and Assessment Administration* (International Association of Assessing Officers (IAAO), 1990, p. 609) caution is advised when the confidence limits are equal to the lowest and highest ratios in the sample. For a 90% median confidence interval, samples of five to seven ratios will have confidence limits equal to the lowest and highest ratios. For a 95% median confidence interval, samples of six to eight ratios will have confidence limits equal the lowest and highest ratios.

D.5 Critical Values For t

Table 19 provides the critical values when calculating mean confidence intervals using a two-tailed test. The steps and assumptions to determine a confidence interval about the mean are included in D.1. of this appendix.

Table 12- Critical Values of t for the two-tailed test

	Confidence level for two-tailed test			
Degrees of Freedom	**70%**	**80%**	**90%**	**95%**
1	1.963	3.078	6.314	12.706
2	1.386	1.886	2.920	4.302
3	1.250	1.638	2.353	3.182
4	1.190	1.533	2.132	2.776
5	1.156	1.476	2.015	2.571
6	1.134	1.440	1.943	2.447
7	1.119	1.415	1.895	2.365
8	1.108	1.397	1.860	2.306
9	1.100	1.383	1.833	2.262
10	1.093	1.372	1.813	2.228
11	1.088	1.363	1.796	2.201
12	1.083	1.356	1.782	2.179
13	1.079	1.350	1.771	2.160
14	1.076	1.345	1.761	2.145
15	1.074	1.341	1.753	2.131
16	1.071	1.337	1.746	2.120
17	1.069	1.333	1.740	2.110
18	1.067	1.330	1.734	2.101
19	1.066	1.328	1.729	2.093
20	1.064	1.325	1.725	2.086
21	1.063	1.323	1.721	2.080
22	1.061	1.321	1.717	2.074
23	1.060	1.319	1.714	2.069
24	1.059	1.318	1.711	2.064
25	1.058	1.316	1.708	2.060
26	1.058	1.315	1.706	2.056
27	1.057	1.314	1.703	2.052
28	1.056	1.313	1.701	2.048
29	1.055	1.311	1.699	2.045
30	1.055	1.310	1.697	2.042
40	1.050	1.303	1.684	2.021
60	1.045	1.296	1.671	2.000
120	1.041	1.289	1.658	1.980
∞	1.036	1.282	1.645	1.960

Note: The table reports positive values only. The rejection region consists of values more extreme than the indicated critical values. For example, the rejection region for a two-tailed test at the 95% confidence level when the degrees of freedom equals 15 is all values greater than 2.131 and less than -2.131. Note that, in this usage, degrees of freedom equal one less than sample size (n-1).

D.6 COD Confidence Intervals

Like measures of central tendency, confidence intervals can be calculated around the Coefficient of Dispersion (COD). The following steps outline the steps to calculate confidence intervals for COD based on the method described in Bonett, Seier (2006).

The Bonett-Seier method creates a COD confidence interval by measuring typical absolute distance from the median, stabilizing it with a log transformation, estimating uncertainty analytically, and transforming back using an exponential transformation, producing reliable intervals for skewed data.

Note: The following steps apply for a 95% confidence interval. Replace 1.96 with 1.645 for a 90% confidence interval. Refer to a t-table for the appropriate t-value for samples less than 30.

1. Calculate the median, mean, and standard deviation from the ratio study data of sample size n. Let $\tilde{x}$ represent the median, $\bar{x}$ represent mean and s represent the standard deviation
2. Convert the standard deviation to the variance, s^2, by squaring the standard deviation
3. Compute the absolute deviation of each ratio from the median ratio
$$ad_i = |x_i - \tilde{x}|$$
4. Compute the mean absolute deviation from the median, $\tau = \sum_{i=1}^{n} ad_i / n$
5. Calculate the mean-median standardized difference, $\delta = (\bar{x} - \tilde{x})/\tau$
6. Calculate the relative dispersion ratio, $\gamma = s^2/\tau^2$
7. Compute the Coefficient of Dispersion, $COD = \tau/\tilde{x}$
8. Rank the ratios in ascending order and determine the positions of the lower variance rank (astar) and the upper variance rank (bstar)
$$astar = Rounded(\left(\frac{(n+1)}{2}\right) - \sqrt{n})$$
$$bstar = (n - astar) + 1$$
9. Identify the ratio values at lower and upper variance ranks. Let ar represent the ratio value at the lower variance rank, star, and br represent the ratio value at upper variance rank, bstar
10. Calculate the log-median variance estimator using the natural logarithm, ln
$$\text{varleta} = \left(\frac{(\ln(\text{ar}) - \ln(\text{br}))}{4}\right)^2$$
11. Calculate the standard error component of the log-median variance estimator
$$\text{se1} = \sqrt{\text{varleta}}$$
12. Calculate the moment-based log scale
$$\text{varltau} = ((\gamma + \delta^2) - 1)/n$$
13. Calculate the standard error component for log dispersion
$$\text{se2} = \sqrt{\text{varltau}}$$
14. Determine the log scale covariance term
$$\text{covltle} = (\delta \times \sqrt{\text{varleta}})/\sqrt{n}$$
15. Calculate the variance calibration factor
$$k = \sqrt{(\text{varleta} + \text{varltau} - (2 \text{ x covltle}))}/(\text{se1} + \text{se2})$$
16. Determine the rank positions for a 95% confidence interval
$$CI_{Rank_{Low}} = \text{Rounded}(((n+1)/2 - (k \text{ x } 1.96 \text{ x } \sqrt{n/4})$$
$$CI_{Rank_{High}} = \left(n - CI_{Rank_{Low}}\right) + 1$$
17. Identify the ratio values located at positions calculated in Step 16. Let $\text{Ratio}_{\text{Rank Low}}$ represent the lower rank ratio and let $\text{Ratio}_{\text{Rank High}}$ represent the higher rank ratio
18. Determine the lower and upper log scale dispersion limits
$$L1 = \ln((n/(n-1)) \times \tau - (k \times 1.96 \times \text{se2}))$$
$$U1 = \ln((n/(n-1)) \times \tau + (k \times 1.96 \times \text{se2}))$$
19. Determine the lower and upper confidence limits for a 95% confidence interval as follows:
$$COD_{Low} = \exp(L1 - \ln(\text{Ratio}_{\text{Rank High}}))$$
$$COD_{High} = \exp(U1 - \ln(\text{Ratio}_{\text{Rank Low}}))$$

D.7 COD Confidence Interval Sample Calculation

This following table provides a summary of the calculations for a 90% COD Confidence Interval based on the data from Table 1 in Section 6.1.

Table 20 – 90% COD Confidence Interval Sample Calculation

Step	Description	Value
	Select Required Confidence Interval	90%
	Corresponding t-value or z-value	1.645
1a	Median ($\tilde{x}$)	0.86778
1b	Mean ($\bar{x}$)	0.89380
1c	Standard Deviation (s)	0.25822
1d	Sample size (n)	54
2	Variance (s^2)	0.06668
3	Absolute Deviation from the Median for individual ratio (ad_i)	$\lvert x_i - \tilde{x} \rvert$
4	Mean Absolute Deviation from the Median (tau (τ))	0.21392
5	Mean-Median standardized difference (delta (δ))	0.12167
6	Relative Dispersion Ratio (gamma (γ))	1.45706
7	Coefficient of Dispersion (COD)	0.24651
8a	Variance_rank_low	20
8b	Variance_rank_high	35
9a	astar ratio value	0.755
9b	bstar ratio value	1.000
10	Log-median variance estimator (varleta)	0.00492
11	Log-median standard error (SE) component (se1)	0.07014
12	Moment-based log-scale variance (varltau)	0.00874
13	Log_dispersion standard error (SE) component (se2)	0.09348
14	Log-scale covariance term (covltle)	0.00116
15	Variance calibration factor (k)	0.65070
16a	CI_rank_low	24
16b	CI_rank_high	31
17a	Log (Ratio_Rank_low)	-0.17229
17b	Log (Ratio_Rank_high)	-0.10384
18a	Lower log-scale dispersion limit (L1)	-1.62354
18b	Upper log-scale dispersion limit (U1)	-1.42342
19a	COD Lower	0.21878
19b	COD Upper	0.28618

D.8 Simple Alternatives to Bootstrap and Formulaic COD Confidence Interval Methods

Bootstrap and formulaic methods are often used to construct confidence intervals for nonparametric statistics like COD because they make few assumptions about the underlying data distribution. However, they require suitable statistical software and may not be practical for many practitioners. The following approach provides a simple alternative that can be implemented with a handheld calculator, spreadsheet, or basic statistical package while still providing a statistically grounded test of whether an observed COD exceeds applicable standards. This approach assumes a normal distribution of ratios and is less precise otherwise.

The formula below generates a "tolerance factor" (TF), similar to those used in quality control and precision testing, that can be multiplied by the maximum allowable COD under IAAO standards to determine whether an observed COD is statistically inconsistent with compliance at a specified confidence level:

$$TF = \text{SQRT}\ (\chi^2 / (n\text{-}1))$$

where:

- x^2 is the chi-square value for a one-tailed test
- n is the sample size

Critical x^2 values represent cutoff values that would be exceeded in only a small percentage (e.g., 5%) of samples of the same size if the COD were truly within the allowable limit. Critical values can be looked up in statistical tables or calculated by simple functions. In Excel, the function is:

CHISQ.INV.RT(sig,df)

For example:

Sig	N	χ2 Formula	χ2 Value
0.05	10	=CHISQ.INV.RT(A2,B2-1)	16.92
0.05	20	=CHISQ.INV.RT(A3,B3-1)	30.14
0.05	30	=CHISQ.INV.RT(A4,B4-1)	42.56

The user, however, need not calculate the critical value separately but can instead calculate the tolerance factor directly:

TF = SQRT(CHISQ.INV.RT(sig, df) / (n − 1))

As another example, in SPSS the equivalent formula is:

TF = SQRT(IDF.CHISQ(sig, df) / (n − 1))

where *sig* is the desired significance level (for example, 0.05 for 95 percent confidence).

TF values can then be multiplied by the maximum allowable COD to determine whether, given the sample size and desired level of confidence, the null hypothesis that the population COD meets applicable standards can be rejected. Assume the maximum allowable COD under IAAO standards is 20. For a sample size of 30 and 95 percent confidence:

$$TF = \text{SQRT}(42.56/29) = 1.21$$

Thus, any related COD exceeding 1.21 x 20 = 24.20 can be deemed out of compliance. A COD below 24.20 indicated that the null hypotheses of compliance cannot be rejected.

Decision Threshold: Cutoff COD = TF × Allowable COD

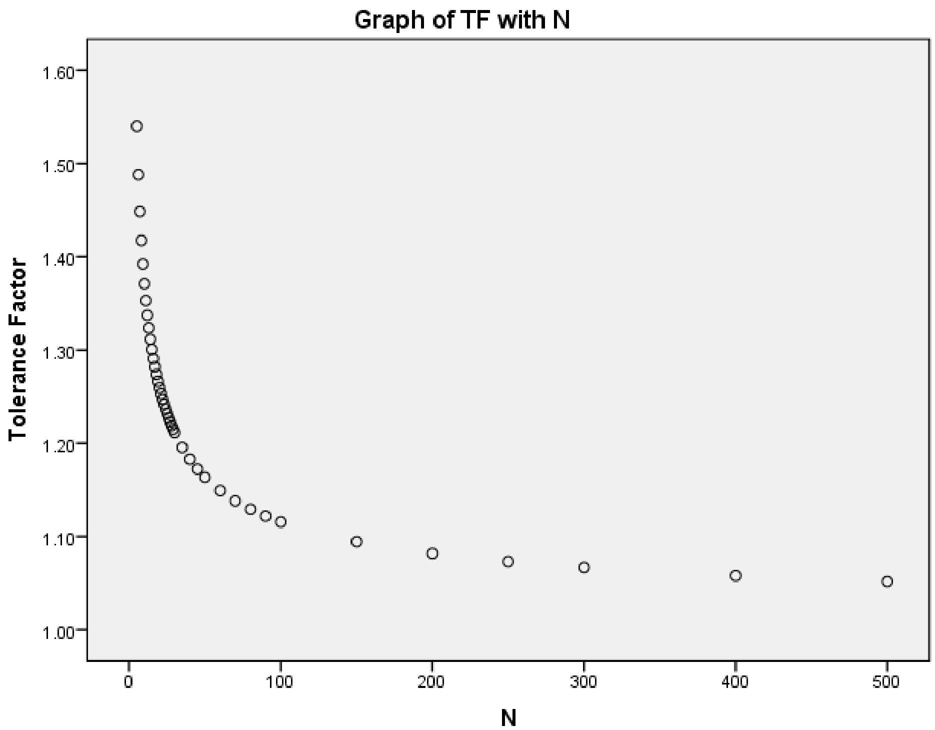
If Study COD ≤ Cutoff COD → **Standard not rejected**

Study COD > Cutoff COD → **Standard rejected**

In summary, this approach tests the hypothesis that the population COD does not exceed a specified cutoff value. It relies on the statistical relationship between sample variance and the chi-square distribution, under which the sampling variability of dispersion statistics is proportional to the square root of sample size. Further, the approach has the advantage that poor CODs – which increase confidence intervals – will not increase tolerance factors, so that poor dispersion is less self-justifying. (For a more detailed explanation of the statistical theory underlying the approach and its application to assessment ratios, see Gloudemans, "Confidence Intervals for the COD: Limitations and Solutions." Property Tax Journal (Nov/Dec 2001).)

The following table, Table 21, contains tolerance factors and cutoff values for various COD standards and sample sizes based on a 95% confidence level. A calculated value below the cutoff values indicates that the null hypotheses of compliance cannot be rejected. Values above indicate that compliance can be rejected and that the sample does not meet the standard. Similar tolerance factors can be used to test a sample COD that is very low and suggests sales chasing.

Notice that tolerance factors decrease as sample size increases, consistent with the statistical principle noted above that sampling precision improves with the square root of sample size. For example, the square root of 25 is 5, while the square root of 100 is 10 – twice as large. Similarly, as the table shows, the tolerance at a sample size of 25 is 0.232, while the tolerance at a sample size of 100 is 0.116 (half as large), yielding approximately twice the precision or reduction in sampling error. The relationship is summarized in the graph below.



The following table provides tolerance factors for various sample sizes and COD standards.

Table 21 – Tolerance Factors and Maximum Allowable CODs

Tolerance Factors and Maximum Allowable CODs					
		-----COD Standard-----			
N	**TF**	**10.00**	**15.00**	**20.00**	**25.00**
5	1.540	15.40	23.10	30.80	38.50
6	1.488	14.88	22.32	29.76	37.20
7	1.449	14.49	21.73	28.97	36.22
8	1.418	14.18	21.26	28.35	35.44
9	1.392	13.92	20.88	27.85	34.81
10	1.371	13.71	20.57	27.42	34.28
11	1.353	13.53	20.30	27.06	33.83
12	1.337	13.37	20.06	26.75	33.44
13	1.324	13.24	19.86	26.47	33.09
14	1.312	13.12	19.67	26.23	32.79
15	1.301	13.01	19.51	26.01	32.52
16	1.291	12.91	19.36	25.82	32.27
17	1.282	12.82	19.23	25.64	32.05
18	1.274	12.74	19.11	25.48	31.85
19	1.266	12.66	19.00	25.33	31.66

Tolerance Factors and Maximum Allowable CODs					
		-----COD Standard-----			
N	TF	10.00	15.00	20.00	25.00
20	1.260	12.60	18.89	25.19	31.49
21	1.253	12.53	18.80	25.06	31.33
22	1.247	12.47	18.71	24.95	31.18
23	1.242	12.42	18.63	24.84	31.04
24	1.237	12.37	18.55	24.73	30.92
25	1.232	12.32	18.48	24.64	30.79
26	1.227	12.27	18.41	24.54	30.68
27	1.223	12.23	18.34	24.46	30.57
28	1.219	12.19	18.28	24.38	30.47
29	1.215	12.15	18.23	24.30	30.38
30	1.211	12.11	18.17	24.23	30.28
35	1.196	11.96	17.93	23.91	29.89
40	1.183	11.83	17.74	23.66	29.57
45	1.172	11.72	17.59	23.45	29.31
50	1.164	11.64	17.45	23.27	29.09
60	1.149	11.49	17.24	22.99	28.73
70	1.138	11.38	17.07	22.76	28.46
80	1.129	11.29	16.94	22.59	28.23
90	1.122	11.22	16.83	22.44	28.05
100	1.116	11.16	16.73	22.31	27.89
150	1.094	10.94	16.42	21.89	27.36
200	1.082	10.82	16.23	21.64	27.05
250	1.073	10.73	16.10	21.46	26.83
300	1.067	10.67	16.00	21.34	26.67
400	1.058	10.58	15.87	21.16	26.45
500	1.052	10.52	15.78	21.04	26.30
750	1.042	10.42	15.64	20.85	26.06
1,000	1.037	10.37	15.55	20.73	25.92

E. VERTICAL EQUITY INDICATOR (VEI)

E.1 Calculation of the VEI

This test is based on the concept found elsewhere in this standard that there should not be a provable difference in valuation level between property groups of greater than 10%. In this case, the test compares value groups and should be conducted as follows:

Step 1: Compute a ratio for each observation by dividing the valuation estimate by its sale price (or time adjusted sale price).

Step 2: Compute a market value proxy for each observation that gives equal weight to sale price (SP) or time adjusted sale price (TASP) and the estimated assessed value (AV).

Proxy = (0.50*SP) + (AV/Median Ratio)

Replace SP with TASP if using time adjusted sale price

Note: When vertical equity detection methods rely solely on the sale price (or third- party estimate of value in the case of an appraisal ratio study) as the market value proxy, results tend to be biased toward regressivity. Likewise, when vertical inequity detection methods rely solely on valuation agency estimates of value as the market value proxy, results tend to be biased toward progressivity. This bias can be reduced by using a market value proxy calculated from the average of the valuation agency value estimate and the sale price, or by using a valid market value proxy not based on either the valuation agency value estimate or the sale price at all.

Step 3: Using the results from steps 1 and 2, sort the market value proxy in ascending order and split the corresponding ratios into percentile groups as outlined in the following table.

Number of Observations	Number of Percentile Groups (PG)
Less than 20	N/A. At least 10 sales per group are required for comparison.
20 – 50	2 (Halves)
51 – 500	4 (Quartiles)
Equal to or greater than 501	10 (Deciles)

For more information on percentile groupings, please see Appendix E, Section E.3.

Step 4: For each percentile group as determined in Step 3, compute the median ratio and the 90% confidence interval.

Note: The formulas for calculating median confidence intervals can be found in Appendix D Section D.2 or (Gloudemans & Almy, 2011 p.365)

Step 5: Calculate the Vertical Equity Indicator (VEI) point estimate as follows:

VEI = 100 * ((MEDIAN Last PG - MEDIAN First PG) / Sample MEDIAN)

The VEI point estimate is calculated by subtracting the first percentile group median from the last percentile group median. Then divide by the overall sample median ratio and multiply by 100.

A negative VEI point estimate indicates a regressive tendency, whereas a positive VEI point estimate indicates a progressive tendency. If the VEI point estimate is within the acceptable range of +/- 10%, then we immediately fail to reject the null hypothesis that the degree of vertical inequity is acceptable.
If the VEI point estimate is outside the acceptable range of +/-10%, then a statistical test must be conducted to determine if there is significant statistical evidence to conclude that the VEI is outside the acceptable range for the population. Proceed to Step 6.

Step 6: If the VEI point estimate lies outside the acceptable range, one recommended method for testing for statistical significance is to compare the confidence interval boundaries for each median. Under this method, first evaluate the upper and lower limits of median confidence intervals for the first and last percentile groups. If the intervals overlap, the null hypothesis immediately fails to reject, and the degree of vertical inequity is within acceptable limits. In other words, the degree of vertical inequity is acceptable. If the confidence intervals do not overlap, proceed to Step 7.

Step 7: If the median confidence intervals do not overlap, calculate the difference between the lower confidence limit for the percentile group with the highest median and the upper confidence limit for the percentile group with the lowest median, scaled by the sample median ratio and multiply by 100 as follows:

VEI Significance = 100 * (Lower CI Limit for PG with the Highest Median - Upper CI Limit for PG with the Lowest Median) / Sample MEDIAN

If the result is greater than 10%, then the test indicates that the difference between the two medians is statistically significantly greater than 10% and thus one can then reject the null hypothesis and conclude there is statistical evidence that the level of vertical inequity present is outside of acceptable limits. In other words, the level of vertical inequity present is not acceptable. If the result is less than 10%, then we fail to reject the null hypothesis that the level of vertical inequity present is within acceptable limits. In other words, the level vertical inequity present is acceptable.

E.2 Further Analysis

It is also highly recommended to graph all calculations for VEI to visualize the relationship between the medians of each percentile group, the accompanying median confidence intervals, and the overall sample median. When all this information is compiled, it can illustrate the presence or absence of vertical equity for a sample.

In situations where the sample has more than two percentile groups, the VEI calculations can also be run on the highest and lowest groups or any other combinations to further evaluate the presence of vertical equity. One should be aware of the statistical considerations surrounding the conduction of multiple comparisons. Vertical inequities can be presented in many shapes so it may not always be the first and last percentile groups which are the highest and lowest medians. If an analyst desires a more comprehensive picture of mass valuation performance in each percentile group, then they could calculate the Coefficient of Dispersion as well for each group

E.3 Percentile Rank Methodologies

According to the United States Department of Commerce's National Institute of Standards and Technology (NIST), there are nine methods to calculate a percentile rank for a set of numbers (https://www.itl.nist.gov/div898/handbook/prc/section2/prc262.htm) based on a paper written by Hyndman and Fan (1996).

Many statistical software packages select either Method 6 (R6) or Method 7 (R7) as the default methodology, although both methods are available. If there are an equal number of observations in each percentile group ranking, R6 and R7 should provide the same percentile ranking. However, if there an unequal number of observations, R6 and R7 may produce slightly different percentile group rankings and may produce different calculations as part of the Vertical Equity Indicator (VEI) analysis.

Method R6

The R6 method uses the following formula to determine the position h_i for each required percentile, P_i:

$$\lfloor h_i \rfloor = \left(\frac{P_i}{100}\right) \times (N + 1)$$

where N is the number of sales ratios in the sample

The floor brackets indicate that h_i is rounded down to the nearest integer as part of the calculation.

Method R7

The R7 method uses the following formula to determine the position h_i for each required percentile, P_i:

$$\lfloor h_i \rfloor = \left(\left(\frac{P_i}{100}\right) \times (N-1)\right) + 1$$

where N is the number of sales ratios in the sample

The floor brackets indicate that h_i is rounded down to the nearest integer as part of the calculation.

Tables 22 and 23 illustrate the differences between R6 and R7 using the data from Table 1 in Section 6.1. Note that the number of observations in each group in both tables is determined as follows:

$$OC_i = \lfloor h_i \rfloor - \lfloor h_{i-1} \rfloor$$

Table 22 – R6 Percentile Rank Groupings using Section 6.1 Table 1 Data

Percentile Group (PG)	P_i	h_i	$\lfloor h_i \rfloor$	Observed Count (OC)
1	25	13.75	13	13
2	50	27	27	14
3	75	41.25	41	14
4	100	54	54	13

N = 54

Table 23 – R7 Percentile Rank Groupings using Section 6.1Table 1 Data

Percentile Group (PG)	P_i	h_i	$\lfloor h_i \rfloor$	Observed Count (OC)
1	25	14.25	14	14
2	50	27	27	13
3	75	40.75	40	13
4	100	54	54	14

N = 54

Table 24 summarizes the available options across some common statistical software and spreadsheet packages.

Software Package	Default Methodology	Function	Alternate Methodology
SPSS	R6	Rank Cases	Write custom syntax to apply R7
R	R7	quantile	Set type = 6 in quantile to apply R6
Excel	R7	PERCENTRANK	N/A
Excel 2010 or later	R7	PERCENTRANK.INC	Use PERCENTRANK.EXC to apply R6
NCSS	R6	Descriptive Statistics	Set Percentile Type to Rank = p(n-1) +1 to apply R7

Software Package	Default Methodology	Function	Alternate Methodology
SAS	R7	PROC UNIVARATE	Set PCTLDEF = 4 to apply R6
Python	R7	numPy	Use sciPy to apply R6

E.4 Sample VEI Calculation

Using the data from Table 1in Section 6.1, the sales ratio and value proxy have been calculated as outlined in Steps 1 and 2 from E.1. Given that there are 54 observations, the data is divided into four groups, as outlined in Step 3 from Section E.1, based on the value proxy.

The overall sample median is 0.8678 and the 90% confidence intervals for the four quartiles are as follows based on the R7 percentile method:

Table 25 – Percentile Group Statistics for VEI using R7 Method

PG	Sales	Med Proxy Value	Median A/S Ratio	Lower CI	Upper CI
Quartile 1	14	$302,709	0.7280	0.6429	0.8476
Quartile 2	13	$436,450	0.8620	0.7867	1.0412
Quartile 3	13	$571.061	0.7537	0.5937	0.9525
Quartile 4	14	$902,512	1.1491	1.0000	1.2476

Note: Because the sales sample is not divisible by 4, there are a different number of sales in two of the quartiles.

To calculate the VEI, we determine that Quartile 1 and Quartile 4 are the first and last percentile groups, such that:

VEI: 100 * ((1.1491 – 0.7280) / 0.8678) = 100 * 0.4853 = 48.53%

Given that the VEI point estimate of 48.53% is greater than our +/- 10% requirement, an evaluation of the confidence interval limits is required.

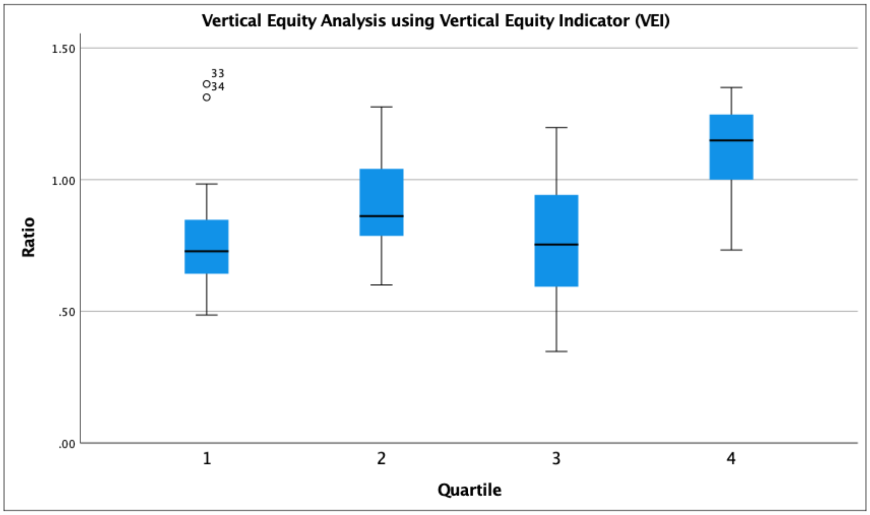
The confidence Intervals for Quartile 1 and Quartile 4 do not overlap, seeing as the lower is from 0.6429 to 0.8476 and the upper is from 1.0000 to 1.2476. Since the confidence intervals (CIs) do not overlap, the VEI Significance must be calculated using the closest levels of these two confidence intervals - 1.000 and 0.8476:

VEI Significance: 100 * (1.0000 - 0.8476) / 0.8678 = 100 * 0.1757 = 17.57%

Since the result of 17.57% is outside the 10% threshold, the requirement of VEI Significance has not been met and the null hypothesis is rejected that the level of vertical inequity present is acceptable. In this instance, since the VEI point estimate is positive, the sales ratios indicate a progressive tendency.

Additionally, plotting the group medians can provide a highly informative visualization of the nature and extent of the vertical inequity present.

Table 26 – Boxplot by Quartile Grouping for VEI



Note that the VEI calculations differ slightly when the R6 percentile method is used on the same sample. Table 27 provides the summary and VEI calculations using the R6 method.

Table 27 – Percentile Group Statistics for VEI Using R6 Method

PG	Sales	Med Proxy Value	Median A/S Ratio	Lower CI	Upper CI
Quartile 1	13	$300,776	0.7172	0.6429	0.8476
Quartile 2	14	$434,726	0.8588	0.7388	1.0412
Quartile 3	14	$584,539	0.7545	0.5937	1.0000
Quartile 4	13	$903,995	1.1635	1.0000	1.2476

VEI = 100*((1.1635 – 0.7172)/0.8678) = 51.43%
VEI Significance = 100*((1.0000-0.8476)/0.8678) = 17.57%
Conclusion: Progressivity

In this instance, the VEI point estimates between both percentile methods differ, but the significance level and conclusion of progressivity are consistent between both methods.

E.5 Summary

The following table summarizes vertical equity outcomes identified in VEI testing:

Table 28 – VEI Outcome Summary Table

Scenario	Result	Outcome
1	VEI Point Estimate within +/-10%	Acceptable
2	VEI Point Estimate outside +/-10% but CIs overlap	Acceptable
3	VEI Point Estimate less than -10% and VEI Significance is less than or equal to 10%	Acceptable

Scenario	Result	Outcome
4	VEI Point Estimate less than -10% and VEI Significance is greater than 10%	Unacceptable Regressivity
5	VEI Point Estimate greater than 10% and VEI Significance is greater than 10%	Unacceptable Progressivity

F. ADDITIONAL TESTS FOR VERTICAL EQUITY

This appendix presents additional tests considered in the IAAO Statistical Tools and Measures Task Force[1]. While the primary measure for vertical equity in the standard is now the Vertical Equity Indicator (VEI) discussed in Section 8.2, there are many tests available to test for vertical equity, each with their strengths and weaknesses. This appendix provides a look at some of these vertical equity tests, however it should by no means be considered a comprehensive list particularly as vertical equity research reaches new conclusions and findings. It is highly recommended that multiple vertical equity tests be run as they may provide additional insights into potential valuation bias.

The following additional methods are summarized in this appendix:

- **Gini Indices** – A set of measures adapted from inequality studies to assess vertical equity, including the Kakwani Index and the Modified Kakwani Index (MKI).
- **Instrumental Variable (IV) Regression** – A technique that directly accounts for measurement errors in market value proxies to assess valuation bias.
- **Mann-Whitney** and **Kruskal-Wallis** – A method that compares valuation levels across percentile groups.
- **Price-Related Bias (PRB)** – A regression-based measure that quantifies the percentage change in assessment ratios relative to changes in value.
- **Price-Related Differential (PRD)** – A simpler index that compares the mean and weighted mean ratios to detect regressivity or progressivity in assessments.
- **Spearman Rank** – A method which calculates the correlation between the rank of valuations and market value proxies.

Gini Indices	
Description	The Gini Coefficient is the basis of the Gini Indices, further discussed in the Kakwani Index (KI) and Modified Kakwani Index (MKI). *Kakwani Index:* Measures the difference between the Gini coefficient of price and the Gini coefficient of assessments, conditional on increasing price levels. This conditional Gini is also called the Concentration Index (CI) since it measures the spread of assessments conditioned on each price level. *Modified Kakwani Index:* Modifies the KI index by taking the ratio, rather than the difference, of the CI of assessments and the Gini coefficient of price. Since it is a ratio, equity is around 1 rather than zero.
Interpretation	The Gini coefficient ranges from 0, indicating perfect equality (where everyone receives an equal share), to 1, perfect inequality (where only one recipient or group of recipients receives all the shares). *Kakwani Index:* A KI equal to zero denotes vertical equity (since the CI of assessments and the gini of prices are the same). A KI less than zero indicates regressive assessments and conversely a KI greater than zero indicates progressive assessments. For lower variability strata, some consider a KI between -0.01 and 0.01 to be optimal. For higher variability or small-sized strata, some consider a KI between -0.02 and 0.02 to be optimal. *Modified Kakwani Index:* An MKI equal to 1 denotes vertical equity (since the assessments and prices have the same Gini measurements). An MKI less than one indicates regressive assessments and conversely an MKI greater than one indicates progressive assessments. For lower variability strata, some consider an MKI between 0.95 and 1.05 to be optimal. For higher variability or small-sized strata, some consider an MKI between 0.90 to 1.10 to be optimal.
Pros	• Naturally has graphical qualities • Allows for a more complete understanding of the nature of any vertical inequity
Cons	• More complicated to calculate and conceptualize • Uses sale price as the market value proxy so can be prone to regressive bias if measurement error in sale prices is sizable enough to significantly reorder the sale price ranks
Additional Resources	Quintos, C. (2020). A Gini measure for vertical equity in property assessments. *Journal of Property Tax Assessment & Administration*, 17(2). Quintos, C. (2021). A Gini decomposition of the sources of inequality in property assessments. *Journal of Property Tax Assessment & Administration*, 18(2).

Instrumental Variable (IV) Regression	
Description	Introduced by Clapp (1990) as a solution to the measurement error problem in prices, this technique accounts for measurement errors in prices to assess valuation bias by creating two regressions.

Instrumental Variable (IV) Regression	
Interpretation	Vertical equity is indicated by the slope coefficient being equal to one in the second stage regression. Regressivity occurs when the slope coefficient is greater than 1; conversely, progressivity occurs when the slope coefficient is less than 1.
Pros	• Mitigates measurement error bias by construction
Cons	• The construction of the instrument variable causes some data loss • More complicated to calculate and conceptualize • Does not allow for an understanding of more complicated vertical equity patterns
Additional Resources	Clapp, John M, 1990. "A New Test for Equitable Real Estate Tax Assessment," The Journal of Real Estate Finance and Economics, Springer, vol. 3(3), pages 233-249, September.

Mann-Whitney	
Description	A nonparametric test of two groups to see if they are being valued at the same percentage of market value. The individual ratios in the two groups are ranked from smallest to largest. The test evaluates if the average ranks assigned to ratios from the two groups are approximately equal.
Interpretation	When the average ranks of the assigned ratios are approximately equal, we accept the null hypothesis that the two groups are appraised at equal percentages of market value. When they are not approximately equal, we reject the null hypothesis and accept the alternative hypothesis that the two groups are not being valued at the same percentage of market value.
Pros	• Easier to understand and calculate • May be useful for smaller sample sizes
Cons	• Only compares two groups so it cannot say much about the nature of the vertical inequity present or the value of the ratio at the extreme ends of the value spectrum
Additional Resources	Fundamentals of Mass Appraisal 380-383

Kruskal-Wallis	
Description	A nonparametric test of three or more groups to see if they are being valued at the same percentage of market value. Like the Mann-Whitney Test, ratios are ranked in ascending order. Using a predetermined confidence interval, such as 95%, the test evaluates whether the differences in the average ranks of the groups are sufficiently large enough to indicate systematic differences in the percentage of market value.

Kruskal-Wallis	
Interpretation	The Kruskal-Wallis Test is the Mann-Whitney Test with more than two groups.
Pros	• Easier to understand and calculate
Cons	• Must ensure that groups are constructed with an adequate sample size • Depending on the number of groups, it may not be able to capture to complete nature of the vertical inequity present
Additional Resources	Fundamentals of Mass Appraisal 383-384

Price-Related Bias (PRB)	
Description	A regression-based test for vertical equity which quantifies the relationship between assessment-sales ratios (ASR) and value in percentage terms by regressing percentage difference from the median ratio on percentage differences in value.
Interpretation	A measure of 0 indicates there is neither regressivity or progressivity present in the sample.
Pros	• Less sensitive to market value proxy outliers than the PRD • The method permits a meaningful interpretation of its coefficient • The method mitigates measurement error bias through the construction of the market value proxy
Cons	• The method assumes a linear relationship between the independent and dependent variables and its coefficient can be misinterpreted should that not be the case • The method may not permit a complete understanding of the nature of any vertical inequity present, only its general direction
Additional Resources	Fundamentals of Mass Appraisal 409-413

Price-Related Differential (PRD)	
Description	An Index statistic which evaluates vertical equity by comparing the mean and weighted mean ratios
Interpretation	A measure of 1.00 indicates there is neither regressivity or progressivity present in the sample. Measures considerably above 1.00 tend to indicate assessment regressivity while measures below 1.00 suggest assessment progressivity.

Price-Related Differential (PRD)	
Strengths	• Easy to compute and interpret
Limitations	• When the weighted mean is heavily influenced by several extreme sales prices, the PRD may not be a sufficiently reliable measure of vertical inequities. • When samples are very large, the PRD may be too insensitive to show small pockets in which there is significant vertical inequity. • The method only yields the general direction of any vertical inequity present, not a complete picture. • Can be biased toward regressivity. • Resampling based methods must be used for confidence intervals and statistical inference.
Additional Resources	Fundamentals of Mass Appraisal 229-231

Spearman Rank	
Description	A nonparametric test in which sales prices and assessment-sales ratios (ASR) are ranked from smallest to largest and correlates the two. This test evaluates the significance of the correlation.
Interpretation	If sales prices and ASR are unrelated, there should be no significant correlation between the two ranks. A negative correlation suggests assessment regressivity while a positive correlation suggests assessment progressivity.
Strengths	• The method has an interpretable result
Limitations	• The method only yields the general direction of any vertical inequity present, not a complete picture. • Uses sale price as the market value proxy so can be prone to regressive bias if measurement error in sale prices is sizable enough to significantly reorder the sale price ranks.
Additional Resources	Fundamentals of Mass Appraisal 387-389

G. LEGAL ASPECTS OF RATIO STUDIES

Property taxation is governed by federal, state, and provincial constitutions, statutes, and administrative rules or regulations, many of which require uniform treatment of property taxpayers. Ratio studies play an important role in judging whether uniformity requirements are met. Relevant Canadian Federal statutes based on the Constitution Acts of 1867-1975 provide that municipal councils cannot discriminate between taxpayers of the same class within municipalities.

Relevant United States federal provisions include the Bill of Rights, the commerce clause of the United States Constitution, the Fourteenth Amendment, and the Tax Injunction Act (28 U.S.C. § 1341). Together they guarantee basic protections and due process while still granting states the authority to classify property and grant reasonable exemptions. Many constitutions have clauses that require uniformity in the assessment and taxation of property, although some jurisdictions, either by constitution or statute, permit certain differences between classes. State courts often weigh in regarding these issues and, in some cases, may also issue reassessment orders. Ratio studies provide a gauge of whether excessive differences exist, or uniformity requirements are being met, and whether compliance with corrective measures has been achieved.

A key U.S. federal statute relating to ratio studies is the U.S. Railroad Revitalization and Regulatory Reform Act ("4-R Act") of 1976 (49 U.S.C. § 11501). The 4-R Act requires that rail transportation property be assessed for tax purposes at no more than 105 percent of the assessment level of other commercial and industrial property in the same taxing jurisdiction.

A key U.S. federal statute relating to ratio studies is the U.S. Railroad Revitalization and Regulatory Reform Act ("4-R Act") of 1976. The 4-R Act requires states to use ratio studies to test whether commercial and industrial property has been assessed at a level that is more than five percent below the ratio of market value at which railroads were assessed and to grant relief if commercial and industrial property is found to be below 95%. In such cases, as in any ratio study, the purpose of the study must be clearly defined, and the study must be conducted so that it accurately evaluates the issues at hand. Important issues in ratio studies conducted pursuant to the 4-R Act include the proper definition of "other" commercial and industrial property, screening and adjustments to sales data, proper measures of the level of appraisal, statistical significance of results, and the combining and weighting of centrally valued and locally assessed properties.

Title 49 U.S.C. § 11501 provides for remediation for acts that unreasonably burden and discriminate against interstate commerce. Section 11501(b) prohibits States and their subdivisions (1) from assessing rail transportation property at values with a higher ratio of assessed value to true market value than the ratio of assessed value to market value of other commercial and industrial property in the same assessment jurisdiction; (2) from levying or collecting a tax on an assessment that violates (1) above; (3) from levying or collecting a tax on railroad property at a rate that exceeds the rate on commercial and industrial property in the same jurisdiction; and (4) from imposing another tax that discriminates against a rail carrier providing transportation services. Section 11501(c) provides that relief may be granted under this subsection only if the ratio of assessed value to true market value of rail transportation property exceeds by at least 5 percent the ratio of assessed value to true market value of other commercial and industrial property in the same assessment jurisdiction. Notably, once the 5 percent threshold is breached remediation typically requires the level of assessment of rail transportation property to be lowered to match that of the other commercial and industrial property.

The 4-R Act provides that ratio studies be used to measure assessment level differences sufficient to be considered discrimination. In such cases, as in any ratio study, the purpose of the study must be clearly defined, and the study must be conducted so that it accurately evaluates the issues at hand. Important issues in ratio studies conducted pursuant to the 4-R Act include the proper definition of "other" commercial and industrial property, screening and adjustments to sales data, proper measures of the level of appraisal, statistical significance of results, and the combining and weighting of centrally valued and locally assessed properties.

Federal statutes, applicable to air transportation property, motor carriers, and bus lines are outlined in 49 U.S.C. §§14502 and 40116.

DEFINITIONS

Absolute value. The value of a number (or variable) regardless of its sign. For example, 3 and -3 (minus 3) both have an absolute value of 3. The mathematical symbol for absolute value is one vertical bar on each side of the number in question, for example, |3|.
Accuracy. The closeness of a measurement, computation, or estimate to the true, exact, or accepted value. Accuracy also can be expressed as a range about the true value. *See also* **precision** *and* **statistical accuracy**.
Adjusted sale price. The sale price that results from adjustments made to the stated sale price to account for the effects of time, personal property, financing, etc.
Appraisal. The act or process of developing an opinion of value; an opinion of value.
Appraisal date. In ad valorem taxation, the date as defined by law to estimate a property's value. *See also* **assessment date**.
Appraisal ratio. (1) The ratio of the appraised value to an indicator of market value. (2) By extension, an estimated fractional relationship between the appraisals and market values of a group of properties. *See also* **level of appraisal**.
Appraisal ratio study. A ratio study using independent — expert single — property appraisals, and/or sales based on market value as indicators of market value. *See also* **Ratio Study**.
Appraisal-sale price ratio. The ratio of the appraised value to the sale price or adjusted sale price of a property; a simple indication of appraisal accuracy.
Appraised value. The estimate of the value of a property.
Arithmetic mean. A measure of central tendency. The result of adding all the values of a variable and dividing by the number of values. For example, the arithmetic mean of 3, 5, and 10 is 18 divided by 3 or 6. Also called **mean**.
Array. An ordered arrangement of data, such as a listing of sales ratios, in order of magnitude.
Assess. Determining the value of property for ad valorem taxation. *See also* **ad valorem tax** *and* **value**.
Assessed value. The value placed on property subject to taxation at market value or some legally authorized fraction thereof. *See also* **Assess**, **Market Value**, *and* **Value.**
Assessment. (1) In general, the official acts of determining the amount of the tax base. (2) As applied to property taxes, the official act of discovering, listing, and appraising property, whether performed by an assessor, a board of review, or a court. (3) The value placed on property in the course of such act.
Assessment-appraisal ratio. The ratio of the assessed value of a property to an independent appraisal.
Assessment date. The status date for tax purposes. Appraised values reflect the status of the property and any partially completed construction as of this date.
Assessment progressivity (regressivity). An appraisal bias such that high-value properties are appraised higher (or lower) than low-value properties in relation to market values. *See also* **price-related differential (PRD)** *and* **coefficient of price-related bias (PRB)**.
Assessment ratio. (1) The fractional relationship of an assessed value to the market value of the property in question. (2) By extension, the fractional relationship of the total of the assessment roll to the total market value of all taxable property in a jurisdiction. *See also* **level of assessment**.
Assessment-sale price ratio. The ratio of the assessed value to the sale price (or adjusted sale price) of a property.
Assessor. Any property tax official with professional responsibilities.
Average absolute deviation. The arithmetic mean of the absolute deviations of a set of numbers from a measure of central tendency such as the median. The average deviation of the numbers 4, 6, and 10 about their median (6) is (2 + 0 + 4) -:-- 3 = 2. The average deviation is used in computing the coefficient of dispersion (COD).
Bias. The tendency of a statistic to overestimate or underestimate a parameter. Statistics describe a sample. A parameter describes an entire population.
Bootstrap. A computer-intensive method of statistical inference that is based on a repeated resampling of data to provide more information about the population characteristics. The bootstrap is a data-driven procedure that is

particularly useful for confidence interval approximation when no traditional formulas are available or the sample has been drawn from a population that does not conform to the normal distribution.

CAMA. *See* **computer-assisted mass appraisal**.

Central tendency. (1) The tendency of most kinds of data to cluster around some typical or central value, such as the mean or median. (2) By extension, any or all such statistics. Some kinds of data, however, such as the weights of cars and trucks, may cluster about two or more values, and in such circumstances the meaning of central tendency becomes unclear. This may happen in ratio studies in which two or more classes of property are combined.

Class. A set of items defined by common characteristics. (1) In property taxation, property classes such as residential, agricultural, and industrial may be defined. (2) In assessment, building classification systems based on type of building design, quality of construction, or structural type are common. (3) In statistics, a predefined category into which data may be put for further analysis. For example, ratios may be grouped into the following classes: less than 0.500, 0.500 to 0.599, 0.600 to 0.699, and so forth. (4) Class use may be subject to regulation and/or law. (5) May have differing tax classes and tax (millage) rates.

COD. *See* **coefficient of dispersion**.

Coefficient of concentration. The percentage of observations within a specified percentage (ie, 15 percent) of a measure of central tendency. Typically used in assessment ratio studies.

Coefficient of dispersion (COD). The average deviation of a group of numbers from the median expressed as a percentage of the median. In ratio studies, the average percentage deviation from the median ratio. The COD is used throughout the property assessment field as a measure of appraisal uniformity.

Coefficient of price-related bias (PRB). An index of price-related bias obtained by regressing percentage deviations from the median ratio on percentage changes in a value proxy, which is obtained by giving equal weight to assessments and sales prices to minimize measurement biases. Indicates the percentage by which assessment ratios change whenever values are doubled or halved. For example, a PRB of −0.03 means assessment levels decrease by 3 percent when value doubles. The PRB should range between −0.05 and +0.05. PRBs outside the range of -0.10 to +0.10 are considered unacceptable.

Coefficient of variation (COV). A measure of relative variability. It is the ratio of the standard deviation to the mean. The COV is particularly useful when comparing results from two groups. For example, in an assessment ratio study, when comparing jurisdictions with widely varying levels of assessment, the COV provides a common basis for reviewing the relative level of assessment equity among the different jurisdictions.

Computer-assisted mass appraisal (CAMA). A software package used by governmental agencies and assessing offices to establish real and personal property valuations for property tax purposes. It is composed of several applications that systemically value property. Often includes statistical analysis such as multiple regression analysis to assist the appraiser in determining the value of property for property taxation purposes.

Confidence interval. A range of values, calculated from the sample observations, are believed, with a particular probability, to contain the true population parameter (mean, median, COD). The confidence interval is not a measure of precision for the sample statistic or point estimate, but a measure of the precision of the sampling process (*see* **reliability**).

Confidence level. The degree of probability associated with a statistical test or confidence interval, commonly 90, 95, or 99 percent. For example, a 95 percent confidence interval implies that were the estimation process repeated again and again, then 95 percent of the calculated intervals would be expected to contain the true population measure (such as the median, mean, or COD).

Contributory value. The amount a component of a property contributes to the total market value. For improvements, contributory value must be distinguished from costs.

COV. *See* **coefficient of variation**.

Date of sale (date of transfer). The date on which the sale was consummated. This is considered to be the date the deed, or other instrument of transfer, is signed. The date of recording can be used as a proxy if it is not unduly delayed as it would be in a land contract.

Direct equalization. The process of converting ratio study results into adjustment factors (trends) and changing locally determined appraised or assessed values to more nearly reflect market value or the legally required level of assessment. *See also* **equalization** *and* **indirect equalization**.
Dispersion. The degree to which data are distributed either tightly or loosely around a measure of central tendency. Measures of dispersion include the range, average deviation, standard deviation, coefficient of dispersion, and coefficient of variation.
Distribution-free statistics. A set of robust nonparametric methods whose interpretation or reliability does not depend on stringent assumptions about the distribution of the underlying population from which the sample has been drawn. *See also* **parametric statistics**.
Equalization. The process by which an appropriate governmental body attempts to ensure that property under its jurisdiction is assessed at the same assessment ratio or at the ratio or ratios required by law. Equalization can be undertaken at many different levels. Equalization among use classes (such as agricultural and industrial property) can be undertaken at the local level, among properties in a school district and a transportation district; equalization among counties is usually undertaken by the state to ensure that its aid payments are distributed fairly. *See also* **direct equalization** *and* **indirect equalization**.
Exploratory data analysis. That part of statistical practice concerned with reviewing the data set to isolate structures, uncover patterns, or reveal features that may improve the confirmatory analysis.
Fixture. An article of personal property installed or attached to real property in such a manner that it is considered to be a part of the real property. The three tests to determine whether an article is a fixture are typically: 1) Extent of annexation to the real property (cannot be removed without substantial damage to the real property; 2) Extent to which the article is a logical contribution to and enhancement of the real property; and 3) Intention of the parties that the article is part of the real property.
Fractional assessments. A percentage of a full value as prescribed by law.
Frequency distribution. A table or chart showing the number or percentage of observations falling in the boundaries of a given set of classes. Used in ratio studies to summarize the distribution of the individual ratios. *See also* **class** *and* **histogram**.
Histogram. A bar chart or graph of a frequency distribution in which the frequencies of the various classes are indicated by horizontal or vertical bars whose lengths are proportional to the number or percentage of observations in each class.
Hypothesis. A statement in inferential statistics, the truth of which the analyst is interested in determining.
Independent appraisal. An estimate of value using a model different from that used for assessment purposes. Independent appraisals are often used to supplement sales in sales ratio studies or in appraisal ratio studies.
Indirect equalization. The process of computing hypothetical values that represent the oversight agency's best estimate of taxable value, given the legally required level of assessment or market value. Indirect equalization allows proper distribution of intergovernmental transfer payments between state or provincial and local governments despite different levels of appraisal between jurisdictions or property classes. *See also* **equalization** *and* **direct equalization**.
Interquartile range (IQR). The result obtained by subtracting the first quartile from the third quartile. By definition 50 percent of the observations fall within the IQR.
Land contract. A contract for the sale of a property in which the seller retains title until the buyer completes the contracted payments for the property. The contract may be recorded; however, the conveyance of the property is not complete until all contractual obligations are fulfilled, at which time the deed for transfer of ownership may be recorded.
Level of appraisal. The common, or overall, ratio of appraised values to market values. Three concepts are usually of interest: 1) the level required by law; 2) the true or actual level; and 3) the computed level based on a ratio study.
Level of assessment. The common or overall ratio of assessed values to market values. *See also* **level of appraisal**. *Note:* The two terms are sometimes distinguished, but there is no convention determining their meanings when they are. Three concepts are commonly of interest: what the assessment ratio is legally required to be, what the assessment ratio for the population actually is, and what the assessment ratio for the population seems to be, on the

basis of a sample and application of inferential statistics. When level of assessment is distinguished from assessment ratio, *level of assessment* usually means either the legal requirement or the true ratio, and *assessment ratio* usually means the true ratio or the sample statistic.

Margin of error. A measure of the uncertainty associated with statistical estimates of a parameter. It is typically linked to consumer surveys or political poll questions. A margin of error is a key component of a confidence interval. It reports a "plus or minus" percentage or proportion quantity in a confidence interval at a specified level of probability (typically 95 percent). *See also* **confidence interval**.

Market value. A value, stated as an opinion, that presumes the transfer of a property (i.e., a right of ownership or a bundle of such rights), as of a certain date, under specific conditions set forth in the value definition that is identified by the appraiser as applicable in an appraisal.

Market value proxy. A property value estimate that is used as a stand-in (proxy) for the unknown true market value of the property.

Mass appraisal. The process of valuing a universe of properties as of a given date using standard methodology, employing common data, and allowing for statistical testing (see *USPAP*).

Mean. *See* **arithmetic mean**.

Median. A measure of central tendency. The value of the middle item in an uneven number of items arranged or arrayed according to size; the arithmetic average of the two central items in an even number of items similarly arranged.

Median absolute deviation. The median of the absolute deviations from the median. In a symmetrical distribution, the measure approximates one-half the interquartile range (IQR).

Median percent deviation. The median of the absolute percent deviations from the median; calculated by dividing the median absolute deviation by one-hundredth of the median.

Nonparametric statistics. *See* **distribution-free statistics**.

Non-sampling error. The error reflected in ratio study statistics from all sources other than sampling error. While non-sampling error is unavoidable due to the inefficiencies inherent in real property markets, the imperfections of the appraisal process, and the imperfections of conducting ratio studies, all practicable steps must be taken to minimize non-sampling error in ratio studies.

Normal distribution. A theoretical distribution often approximated in real world situations. It is symmetrical and bell-shaped; 68 percent of the observations occur within one standard deviation of the mean and 95 percent within two standard deviations of the mean, and 99 percent are three standard deviations of the mean.

Observation. One recording or occurrence of the value of a variable, for example, one sale ratio among a sample of sales ratios.

Outliers. Observations that have unusual values, that is, differ markedly from a measure of central tendency. Some outliers occur naturally; others are due to data errors.

Overfitting. Where a statistical model contains too many variables relative to the sample size of observations. A good test is to divide the number of observations by the number of independent variables. If the average number of observations per independent variable falls below 30, the risk of overfitting becomes much higher. This leads to poor performance on new data. When a model is overfit, relatively poor results from a ratio study of sales after the time period (relative to the sales used in the model) may be seen. .

Parameter. Numerical descriptive measure of the population, for example, the arithmetic mean or standard deviation. Parameters are generally unknown and estimated from statistics calculated from a sample of the population.

Parametric statistics. A statistic that assumes the sample data come from a population with a normal distribution. *See also* **distribution-free statistics**.

Percentile. The values that divide a set of data into specified percentages when the data are arrayed in ascending order. The tenth percentile includes the lowest 10 percent of the values, the twentieth percentile includes the lowest 20 percent of the values, and so forth.

Personal property. All property not defined as real property. Generally, includes four categories: inventories, fixed assets, leasehold improvements, and intangibles). Tangible property includes most movable items, but can include sales tax, freight, installation costs, and all other costs required to place the item into service. Intangible property is evidence of ownership. Personal property is assessable unless specifically exempted by statute for each state. *Also see* **property**.

Plottage value. The incremental value increase that results when two or more sites are combined to form one larger site with greater utility.

Point estimate. A single numerical value used to estimate a population parameter from a sample. Point estimates are generally constructed to provide the best unbiased estimate of the population parameter consistent with the sample data. However, the point estimate is only an estimate, and is unlikely to have the same value as the population parameter. (*See* **Confidence interval** and **Reliability** for discussion of precision of the sampling process.)

Points. Prepaid interest on a loan; one point is equal to 1 percent of the amount of the loan. It is common to deduct points in advance of the loan, so that an individual pays interest on 100 percent of the loan but gets cash on, say, only 99 percent.

Population. All the items of interest, for example, all the properties in a jurisdiction or neighborhood; all the observations in a data set from which a sample may be drawn.

Precision. The level of detail in which a quantity or value is expressed or represented. It can be characterized as the number of digits used to record a measurement. A high level of represented precision may be used to imply a greater level of accuracy; however, this relationship may not be true. Precision also relates to the quality of an operation or degree of refinement by which results are obtained. A method of measurement is considered precise if repeated measurements yield the same or nearly the same numeric value. *See also* **accuracy** *and* **statistical precision**.

PRB. *See* **coefficient of price-related bias**.

PRD. *See* **price-related differential**.

Price. The amount asked, offered, or paid for a property.

Price-related differential. A statistical measure of vertical property tax equity. The PRD is calculated by dividing the mean ratio by the weighted mean ratio. Price-related differentials above 1.03 tend to indicate assessment regressivity; price-related differentials below 0.98 tend to indicate assessment progressivity.

Primary Valuation Office. The unit of government having initial responsibility for determining the assessed value against which general property taxes are levied by local government and, where applicable, by state government.

Progressivity. *See* **assessment progressivity (regressivity)**.

Property. An aggregate of things or rights to things. These rights are protected by law. There are two basic types of property: real and personal. Real property consists of the interests, benefits, and rights inherent in the ownership of land plus anything permanently attached to the land or legally defined as immovable; the bundle of rights with which ownership of real estate is endowed. To the extent that "real estate" commonly includes land and any permanent improvements, the two terms can be understood to have the same meaning. Also called *realty*. Personal property is defined as those items that generally are movable or all items not specifically defined as real property. Many states include as personal property the costs associated with placing personal property in service, such as sales tax, freight, and installation. Installation items include, but are not limited to, wiring, foundations, hookups, and attachments. Two commonly used tests for distinguishing real and personal property are (1) the intent of the parties and (2) whether the item may be removed from the real estate without damage to either.

Qualified sale. A property transfer that satisfies the conditions of a valid sale and meets all other technical criteria for inclusion in a ratio study sample. If a property has undergone significant changes in physical characteristics, use, or condition in the period between the assessment date and sale date, it would not technically qualify for use in ratio study.

Quartiles. The values that divide a set of data into four equal parts when the data are arrayed in ascending order. The first quartile includes the lowest quarter of the data, the second quartile, the second lowest quarter, and so forth.

Random sample. A sample of n items selected from a population in such a way that each sample of the same size is equally likely. This also includes the case in which each element in the sample has an equal chance of being selected.

Range. (1) The maximum value of a sample minus the minimum value. (2) The difference between the maximum and minimum values that a variable may assume.

Ratio study. A statistical study of the relationship between appraised or assessed values and market values; based on an analysis of the ratio derived by dividing the appraised or assessed values of property by the market values of such property. Sale prices or independent appraisals are used as proxies for market values. Of common interest in ratio studies are the level and uniformity of the appraisals or assessments. *See also* **level of appraisal** *and* **level of assessment**.
Real property. Consists of the interests, benefits, and rights inherent in the ownership of land plus anything permanently attached to the land or legally defined as immovable; the bundle of rights with which ownership of real estate is endowed. To the extent that "real estate" commonly includes land and any permanent improvements, the two terms can be understood to have the same meaning. Also called Realty. *See* **property**.
Regressivity. *See* **assessment progressivity (regressivity)**.
Regressivity index. *See* **price-related differential**.
Reliability. In a sampling process, the extent to which the process yields consistent population estimates. Ratio studies typically are based on samples. Statistics derived from these samples may be more or less likely to reflect the true condition in the population depending on the reliability of the sample. Representativeness, sample size, and sample uniformity all contribute to reliability. Formally, reliability is measured by sampling error or the width of the confidence interval at a specific confidence level relative to the central tendency measure.
Representative sample. A sample of observations from a larger population of observations, such that statistics calculated from the sample can be expected to represent the characteristics of the population being studied.
Sale price. (1) The actual amount of money exchanged for a unit of goods or services, whether or not established in a free and open market. An indicator of market value. (2) Loosely used synonymously with "offering" or "asking price."
Sale ratio. The ratio of an appraisal (or assessed) value to the sale price or adjusted sale price of a property.
Sales chasing. (1) The practice of using the sale of a property to trigger a reappraisal of that property at or near the selling price. If sales with such appraisal adjustments are used in a ratio study, the practice causes invalid uniformity results and causes invalid appraisal level results, unless similar unsold parcels are reappraised by a method that produces an appraisal level for unsold properties equal to the appraisal level of sold properties. (2) By extension, any practice that causes the analyzed sample to misrepresent the assessment performance for the entire population as a result of acts by the assessor's office. A subtle, possibly in- advertent, variety of sales chasing occurs when the recorded property characteristics of sold properties are differentially changed relative to unsold properties. Then the application of a uniform valuation model to all properties results in the recently sold properties being more accurately appraised than the unsold ones.
Sales ratio study. A ratio study that uses sales prices as benchmarks for market values. A relationship between sales prices and value (market value, assessed value, equalized value), that is used to measure the level of appraisal. Used to evaluate the effectiveness of assessment practices, reappraisals, or revaluations.
Sample. A subset containing the characteristics of a larger population. A sample should represent the population as a whole and not reflect any bias toward a specific attribute. In order to achieve an unbiased sample, the selection must be random so each item from the population has an equal chance of inclusion in the sample group.
Sampling error. A statistical error from a sample that does not represent the population.
Scatter diagram or scatter plot. A graphic means of depicting the relationship or correlation between two variables by plotting one variable on the horizontal axis and one variable on the vertical axis. Often in ratio studies it is informative to determine how ratios are related to other variables. A variable of interest is plotted on the horizontal axis and ratios are plotted on the vertical axis.
Significance level. A measure of the probability that an event is attributable to a relationship rather than merely the result of chance.
Skewed. The quality of a frequency distribution that makes it asymmetrical. Distributions with longer tails on the right than on the left are said to be skewed to the right or to be positively skewed. Distributions with longer tails to the left are said to be skewed to the left or to be negatively skewed.
Standard deviation. The statistic calculated from a set of numbers by subtracting the mean from each value and squaring the remainders, adding together all the squares, dividing by the size of the sample less one, and taking the

square root of the result. When the data are normally distributed, the percentage of observations can be calculated within any number of standard deviations of the mean from normal probability tables. When the data are not normally distributed, the standard deviation is less meaningful, and the analyst should proceed cautiously.
Standard error. A measure of the precision of a measure of central tendency; the smaller the standard error, the more reliable the measure of central tendency. Standard errors are used in calculating a confidence interval about the arithmetic mean and the weighted mean. The standard error of the sample mean is the standard deviation divided by the square root of the sample size and can only be estimated, unless the real population parameter is known A common measure of sampling error. The difference between a population parameter and a sample statistic. The standard error of the mean is an estimate of the precision of the sample mean. While the standard error of the mean is the most common type of standard error, standard errors are calculated for other statistics as well. The standard deviation is a descriptive statistic that can be calculated from sample data. In contrast, the standard error is an inferential statistic that can only be estimated, unless the real population parameter is known.
Statistical accuracy. The closeness between the statistical estimate and the true (but unknown) population parameter value it was designed to measure. It is usually characterized in terms of error or the potential significance of error and can be decomposed into sampling error and non-sampling error components. Accuracy can be specified by the level of confidence selected for a statistical test. *See also* **accuracy**.
Statistical precision. How close estimates from different samples are to each other. For example, the standard error is a measure of precision. When the standard error is small, estimates from different samples will be close in value. Precision is inversely related to standard error. When the standard error is small, sample estimates are more precise; when the standard error is large, sample estimates are less precise. *See also* **precision**.
Statistics. Numerical descriptive data calculated from a sample, for example, the median, mean, or COD. Statistics are used to estimate corresponding measures, termed parameters, for the population.
Stratify. To divide, for purposes of analysis, a sample of observations into two or more subsets according to some criterion or set of criteria.
Stratum, strata (pl.). A class or subset that results from stratification.
Time-adjusted sale price. The price at which a property sold adjusted for the effects of price changes reflected in the market between the date of sale and the date of analysis.
Trimmed mean. The arithmetic mean of a data set identified by the proportion of the sample that is trimmed from each end of the ordered array. For example, a 10 percent trimmed mean of a sample of size ten is the average of the eight observations remaining after the largest and smallest observations have been removed.
Value. (1) The relationship between an object desired and a potential owner; the characteristics of scarcity, utility, desirability, and transferability must be present for value to exist. (2) Value may also be described as the present worth of future benefits arising from the ownership of real or personal property. (3) The estimate sought in a valuation. (4) Any number between positive infinity and negative infinity. *See* **market value**.
Variable. An item of observation that can assume various values, for example, square feet, sales prices, or sales ratios. Variables are commonly described by using measures of central tendency and dispersion.
Weighted mean; weighted average. An average in which each value is adjusted by a factor reflecting its relative importance in the whole before the values are summed and divided by their number.
Weighted mean ratio. Sum of the appraised values divided by the sum of the sales prices (or independent estimates of market value), which weights each ratio in proportion to the sale price (or independent estimate of market value).

REFERENCES

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data.* New York: John Wiley & Sons, Inc.

Clapp, J. M. (1989). Sample size in ratio studies: How can "small" be made "large enough". *Property Tax Journal, 8*(3), 211-231.

Cochran, W. G. (1977). *Sampling techniques.* New York: John Wiley & Sons, Inc.

D'Agostino, R. B., & Stephens, M. (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker.

Dornfest, A. S., & Chizewsky, A. N. (2017). Equalization Revisited: Achieving Representativeness with the Weighted Mean Ratio. *Journal of Property Tax Assessment and Administration (JPTAA), 14*(2), 45-58. Retrieved from https://researchexchange.iaao.org/jptaa/vol14/iss2/3/

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Force, I. S. (2023). A review of vertical equity measures in property assessments. *Journal of Property Tax Assessment and Administration (JPTAA), 20*(2), 131. doi:https://doi.org/10.63642/1357-1419.1265

Gloudemans, R. (1999). *Mass appraisal of real property.* Chicago: International Association of Assessing Officers.

Gloudemans, R., & Almy, R. (2011). *Fundamentals of mass appraisal.* Kansas City, Missouri: International Association of Assessing Officers.

Hermans, L., Bidanset, P., Davis, P., & McCord, M. (2022). Using property-level ratio studies for the incorporation of validation models in single-family residential real estate assessment. *Journal of Property Tax Assessment and Administration (JPTAA), 19*(1), 83-103. doi:https://doi.org/10.63642/1357-1419.1245

Hermans, L., McCord, M., Davis, P., & Bidanset, P. (2023). An exploratory approach to composite modelling for real estate assessment and accuracy. *Journal of Property Tax Assessment and Administration (JPTAA), 20*(1), 43-61. doi:https://doi.org/10.63642/1357-1419.1256

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: John Wiley & Sons, Inc.

Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers.* Milwaukee: ASQC Quality Press.

International Association of Assessing Officers (IAAO). (1978). *Improving real property assessment: A reference manual.* Chicago: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (1990). *Property appraisal and assessment administration.* Chicago: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2003). *Assessment administration.* Chicago, Illinois: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2011). *Fundamentals of Mass Appraisal.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2018). *Standard on automated valuation models.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2018). *Standard on valuation of personal property.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2020). *Standard on oversight agency responsibilities.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2020). *Standard on property tax policy.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2020). *Standard on verification and adjustment of sales.* Kansas City, Missouri: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2021). *Standard on Data Quality.* Kansas City, MO: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). (2025). *Standard on mass appraisal of real property.* Kansas City, MO: International Association of Assessing Officers.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Non-sampling error in surveys.* New York: John Wiley & Sons, Inc.

Sherrill, K., & Whorton, Jr, E. (1991). Sample size estimation techniques of the state equalization study of school districts in Texas. *Property Tax Journal, 10*(1), 125-139.

The Appraisal Foundation. (2024 (Updated every two years)). *Uniform standards of professional appraisal practice (USPAP).* Washington, DC: The Appraisal Foundation.

Tomberlin, N. (2001). Trimming outlier ratios in small samples. *Assessment Journal, 8*(4), 29-35.

Wooten, T. (2003). Asking the right questions is the key to a valid ratio study analysis. *Assessment Journal, 10*(4), 97-102.

ADDITIONAL RESOURCES

The Appraisal Foundation. 2024 (updated every two years). *Uniform standards of professional appraisal practice* (USPAP). Washington, DC: The Appraisal Foundation.

Committee on Sales Ratio Data, National Association of Tax Administrators. 1954. Report of the Committee. *Guide for assessment-sales ratio studies*. Chicago: Federation of Tax Administrators.

Benmamoun, M. 2006. Bootstrap confidence intervals and Gloudemans' COD tolerance test using SPSS and Stata. *Journal of Property Tax Assessment and Administration* 3(4):56-61.

Birch, J.W. and M.A. Saunderman. 2000. Optimal Trimming of appraisal-sales ratio studies. *Assessment Journal* 6(4):25-31.

Birch, J.W. and M.A. Saunderman. 1997. Testing for outliers in sales and appraisal data. *Assessment Journal* 4(4):31-42.

Bonett, D.G. and E. Seier. 2006. Confidence intervals for a coefficient of dispersion in nonnormal distributions. *Biometrical Journal*. 48(1):144- 148

Clapp, John. 1990. "A methodology for constructing vacant land price indices." Journal of the American Real Estate and Urban Economics Association 18(3): 274-293

Conover, W. J. 1980. *Practical nonparametric statistics*. New York: John Wiley & Sons.

D'Agostino, R.B., A. Belanger, and R.B. D'Agostino, Jr. 1990. A suggestion for using powerful and informative tests for normality. *The American Statistician* 44(4):316-21.

Denne, R.C. 2011. The PRB and other potential successors to the flawed PRD as a measure of vertical assessment inequity. *Fair & Equitable* 9(11): 3-10.

Dornfest, A.S., A. Chizewsky, and P. Davis. 2004. Alternate methods of addressing ratio study outliers. *Journal of Property Tax Assessment and Administration* 1(4):5-14.

Dornfest, A.S. 1993. Mass appraisal performance evaluation: Strategies for painless implementation. *Assessment Digest* 15(1):2-11.

Dornfest, A. S. 2000. State and provincial ratio study practices: 1997 Survey results. *Assessment Journal* 4(6):23-67.

Dornfest, A.S. 1990. Perspective on ratio studies: The rural state. *Assessment Digest* 12(3):17-21.

Gloudemans, R.J. 1990. Adjusting for time in computer- assisted mass appraisal. *Property Tax Journal* 9(1):83-99.

Gloudemans, R.J. 2000. An empirical evaluation of central tendency measures. *Assessment Journal* 7(1):21-27.

Gloudemans, R.J. 2001. Confidence intervals for the COD: Limitations and solutions. *Assessment Journal* 8(6):23-27.

Gloudemans, R.J. 2011. The coefficient of price-related bias: A measure of vertical equity. *Fair & Equitable* 9(8): 3-8.

Groves, R.M. 2004. *Survey errors and survey costs*. New York: John Wiley & Sons.

Hart, Anna. 2001. Mann-Whitney test is not just a test of medians: Differences in spread can be important. *British Medical Journal* 2001(323):391-393.

International Association of Assessing Officers (IAAO). 2003. *Standard on facilities, computers, equipment and supplies*. Chicago: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). 1977. *Analyzing assessment equity*. Chicago: International Association of Assessing Officers.

International Association of Assessing Officers (IAAO). 2022. *Glossary for property appraisal and assessment*. Chicago: International Association of Assessing Officers. https://www.iaao.org/wp-content/uploads/IAAO-Glossary3rd-Ed_final.pdf

Jacobs, T. 1986. Assessment quality control. *Assessment Digest* 8(4):8-13.

Knight,John R., Thomas J. Miceli, and C.F. Sirmans. 2000. Repair expenses, selling contracts and house prices. *Journal of Real Estate Research* 20(3).

Mendenhall, W., J.E. Reinmuth, and R.J. Beaver. 1993. *Statistics for management and economics*, 7th ed. Belmont, CA: Duxbury Press.

Neave, H. R., and P.L. Worthington. 1988. *Distribution-free tests*. London, England, and Boston, MA: Unwin Hyman.

Neter, J., W. Wasserman, and G.A. Whitmore. 1987. *Applied statistics*. 3rd ed. Boston, MA: Allyn and Bacon, Inc.

Schultz, Ronald J. 1996. The law of the tool: A question of fairness. *Assessment Journal* 3(6):62-70.

Snedecor, G.W., and W.G. Cochran. 1967. *Statistical methods*, 6th ed.Ames Iowa: Iowa State University Press.

Tomberlin, Nancy. 2001. Sales validation from an oversight agency's perspective. *Assessment Journal* 8(6):29-35]

Twark, Richard D., Raymond W. Everly, and Roger H. Downing. 1989. Some insights into understanding assessment uniformity measures: Regressivity and progressivity. *Property Tax Journal*. 8(3):183-91.

Whorton, E.B. Jr. 2003. Should sales ratio studies be used for equalizing state funds for schools? *Assessment Journal*. 10(2):29-40.

Wu, K. and R. Baker, Jr. 2000. Using Tukey line in support of PRD. *Assessment Journal* 7(5):41-50.

Wu, K. and R. Baker, Jr. 2000. Evaluating assessment performance using a composite index: A tool for communication. *Assessment Journal* 7(1):29-33.

Additional readings on ratio studies can be found within the IAAO Library which can be found at http://www.iaao.org. Many websites offer good information on statistics. Because website addresses change frequently, they are not listed here.