Construction of Strata Boundaries in Tax Auditing

Zachary Rhyne[1], Roger Pfaffenberger[1]
[1] Ryan LLC, Three Galleria Tower, 13155 Noel Road, Suite 100, Dallas, TX 75420

Abstract
The cumulative square root of the frequency method is a generally accepted statistical technique used for the construction of strata boundaries in stratified sampling. Many statistical consultants and state and federal taxing and auditing agencies utilize this statistical method originally developed by Dalenius and Hodges (1959). But there is a general lack of guidance on the determination and effects of interval (i.e., class) widths. Dalenius and Hodges originally proposed the application of their method using frequency distributions with class widths of 5 units. In this paper, we present the results of empirical tests to contrast Dalenius' method with different class widths and to other approximate, non-iterative methods using several typical skewed accounting populations.

Key Words: statistical sampling, tax auditing, cumulative square root of the frequency, class width

1. Introduction

The cumulative root frequency method (the "$cum\sqrt{f}$ method") is a general accepted statistical method to construct strata boundaries. The purpose of the $cum\sqrt{f}$ method is to approximate optimal boundaries by minimizing the product of the stratum weight multiplied by the true variance which the method seeks to accomplish by equalizing the $cum\sqrt{f}$ across the strata (Cochran 1977). It has been widely used for decades in audit sampling for the review of documentation associated with monetary amounts. It continues to be among the most prevalent and recommended design methods for consideration in revenue sampling manuals, audit sampling literature, statistical sampling research, and audit and statistical software.[1]

Even with this historical use and the abundance of statistical literature on design methods, there is little guidance on determining the appropriate width of the interval or stated differently the number of class intervals. As Hedlin (2000) explained, there is no existing theory that determines the best interval width. Much of the statistical sampling literature recommends many classes or in other words small interval widths (Cochran 1961; Cochran 1977; Roberts 1978; Hedlin 2000; Hogan 2010). The reasoning behind this recommendation is that with a more refined class interval is a greater potential for precision in the equalization of the $cum\sqrt{f}$ across the classes and thus for optimizing boundaries. But this guidance is general and provides no specific recommendation for determining the number of class intervals or interval widths other than "many" and "narrow."

[1] While we hope this research will be informative to sampling statisticians, the paper has specifically been written for consideration by revenue agents, auditors, tax consultants, and other practitioners that use and encourage sampling.

Notwithstanding the class widths chosen (and thus by association the final strata boundaries), Hedlin (2000) statistically reasoned that in practice "this might not be severe, as the estimator variance regarded as a function of the stratum boundaries is usually flat around its minimum, which makes minor deviations from the minimum negligible."[2]

With this history and guidance, the applied statistician and audit practitioner are often simply told to "determine" the class width with no further recommendations which has given rise to some informed but far from effective sample designs, some haphazard designs with million dollar consequences, and some casual audit sampling practices for implementing the $cum\sqrt{f}$ method. For example, the guidance on using many class intervals has led some auditors to take the approach of extremely narrow interval widths (e.g., $0.01, $1.00) which this paper will explore further.

Dalenius and Hodges (1959) originally proposed the application of their $cum\sqrt{f}$ method using interval widths of 5 units. This specific interval width of 5 has thus been applied formulaically by some in the audit practice creating a legacy technique without consideration to a population's unique characteristics.

Referring to the class widths in the frequency distribution of Table 5A.11 in Cochran's (1977) *Sampling Techniques,* there is no statistical requirement about the class width of 5. It is our opinion that large data files were still being summarized in frequency distributions in 1977 (and often still are for the purpose of producing histograms, etc.) because of the difficulties of working with data populations without a mainframe computer. The intended purpose of a frequency distribution is to summarize large datasets in meaningful ways. The general guidelines were to have between 5 and 20 classes which facilitated the development of histograms and line charts that highlight interesting features of the data. This practice too has found its way into the application of the $cum\sqrt{f}$ method (and is not without its own merits when applied with an understanding of the data populations under consideration).

The generalness of these guidelines has led other audit practitioners to believe there is no optimal class width and divide their populations up into wide intervals to simply allow for a quicker processing of data (and the subsequent stratification) without consideration to the statistical consequences of those widths. While there are other ways in which auditors determine class width (some ways better than others), these examples will suffice to illustrate the need for more detailed guidance.

It is not our purpose to focus on creating optimal bounds (though we expect this study to contribute to that body of literature) but to explore if varying class width has an effect or is negligible and what influence varying class widths have on the accuracy and precision of estimates produced from samples selected from $cum\sqrt{f}$ designs and contrasted to other design methods.

2. Construction of Strata Boundaries in Tax Auditing

[2] Hedlin, D. (2000), "A procedure for stratification by an extended Ekman rule," Journal of Official Statistics, 16, 15-29.

There are several generally accepted statistical sampling techniques for the construction of strata boundaries outside of the $cum\sqrt{f}$ method that are interesting and used by statisticians, but they lack practicality for most practitioners in audit sampling. Some of these methods are less frequently used for several reasons. They are intensive programmatic iterative algorithms, the algorithms have potential execution flaws due to numerical complications, a lack of common audit software support, an absence of practical guidance, etc. As Cochran (1961) concluded on approximately equalizing the standard deviation (S_h) among strata, these kinds of methods could be to intricate to implement in most sampling practices and often remain outside the purview of audit practices that lack strong operation resources and access to deep programming and statistical technical skills.

Within tax auditing, both at the federal and state levels, the most frequently used, software supported, or programmatically designed methods are the $cum\sqrt{f}$ method, the equal $ method (Roberts 1978), the geometric method (Gunning, Horgan, and Yancy 2004), the equalization of the product of the weight and standard deviation, W_hS_h (Cochran 1977), and consideration to the coefficient of variation, CV_h (Cochran, 1961).

2.1 Comparing Stratification Methods

Twelve scenarios of design techniques or derivatives of those methods are explored in this paper to conclude on the effects of varying class widths and comparative performance of those interval widths to other common audit sampling design methods.

The $cum\sqrt{f}$ method at the following seven classes widths: $0.01, $1, $5, $50, $100, $250, $500. These widths consider the narrowest, narrow, moderate, wide, and comparatively width interval widths. The $0.01 width creates a great statistical quality to explore around the empirical cumulative distribution function (and its use is some audit software). Additionally, one other $cum\sqrt{f}$ width interval is considered and determined by a simple formula: median/100. This is of interest as it is dynamic and leverages a population's attributes.

The equal $ amount method is an important technique for its simplicity of use and robust and practical handling of skewed, large dollar account populations. Many audit sampling packages support this technique and it is another historical and prevalent design method in audit and financial sampling.

The geometric method per Gunning and Horgan (2004) frequently creates constant coefficients of variation among strata. This is a great statistical quality and one that Cochran (1961) remarks is frequently associated with optimal strata boundaries. It is also another design method supported by common audit sample packages though it is a comparatively new method. Its weakness, because of the inherit mathematical property of geometric progression, is with populations that have very small monetary amounts that must be sampled (Gunning and Horgan 2004). But this weakness can be advantageous as many audit situations will apply a de minimis threshold and set aside small valued items as being negligible to their review.

The equalization of counts among strata (N_h) is explored as a method that is far less frequently used and far less statistically associated with efficient designs. Thus, it serves as a barometer in this research to the other methods and their derivatives. Finally, the research considers making constant the number of unique dollar amounts among strata. Not a technique the authors are aware of being used but a derivative on Equal (N_h) and one that possess several of the same qualities of the $cum\sqrt{f}$ method using "narrowest" widths without creating empty classes.

Table 1: The Twelve Design Method Scenarios

(cum. √ f): 0.01	(cum. √ f): 1	(cum. √ f): median/100	(cum. √ f): 5	(cum. √ f): 50	(cum. √ f): 100	(cum. √ f): 250	(cum. √ f): 500	Equal $ Amount	Geometric	Equal Nh	Equal Nh per unique $
m01	**m02**	**m03**	**m04**	**m05**	**m06**	**m07**	**m08**	**m09**	**m10**	**m11**	**m12**

3. Populations of Interest and Simulation Methodology

The population data of recorded amounts[3] of auditing, tax, accounting, business, financial, and in general monetary amounts often possess a "lack of symmetry about the population mean" as shown by their frequency distributions. This skewness can be extreme in audit situations with a majority of small and moderate amounts and few very large amounts (Roberts 1978). The tails of the population distributions are heavier than a normal distribution. For our purpose, the recorded amounts are the design (i.e., stratification) variable amounts and not the audited amounts. When graphing the frequency of differences (where differences = recorded amount – audited amount) an examiner regularly discovers high symmetry and a strong concentration of values around zero amount differences.

3.1 Accounting Populations

This research explores generally accepted stratum construction methods on common audit populations and for this purpose nineteen homologous accounting populations were developed paralleling data distributions such as is found in common industries (e.g., Oil & Gas, Financial Services, etc.) and accounting populations (Sales & Use Tax, Meal & Entertainment Expenses, etc.). Population d1 is the same as population d13 as is the population used by Rhyne & Falk (2007) in exploring treatment of negative values in audit populations. When creating the populations, all recorded amounts less than or equal to zero were removed. Table 2 includes the measurements of skewness and kurtosis to evaluate normality along with the standard population statistics. References in Table 2 to detail threshold and detail stratum will be explained in a following section of this paper under sample design.

Table 2: Homologous Accounting Populations Constructed & Reviewed
(Rounding for visual aesthetics and consumability)

										Without Detail Stratum	
Data	Amount	Count	Min	Median	Mean	Standard Deviation	Skewness	Kurtosis	Detail Threshold	Skewness	Kurtosis
d01	$ 74,500,000	20,000	$ 0.4	$ 200	$ 4,000	53,000	60	4,000	$ 50,000	10	40
d02	$ 2,300,000,000	22,000	$ 1.0	$ 300	$ 105,000	1,800,000	40	2,000	$ 7,000,000	20	300
d03	$ 169,900,000	8,000	$ 0.1	$ 4,000	$ 23,000	81,000	10	200	$ 440,000	5	30
d04	$ 98,200,000,000	58,000	$ 0.1	$ 60,000	$ 1,706,000	26,300,000	80	8,000	$ 260,000,000	20	350
d05	$ 1,400,000	500	$ 1,000.0	$ 1,800	$ 3,000	4,000	10	100	$ 10,000	2	4
d06	$ 12,000,000	3,000	$ 1.4	$ 1,800	$ 4,000	6,000	3	10	$ 30,000	2	5
d07	$ 7,200,000,000	10,000	$ 2.4	$ 55,000	$ 748,000	3,800,000	10	200	$ 6,000,000	4	10
d08	$ 500,000	500	$ 0.1	$ 300	$ 1,000	5,000	10	100	$ 1,900	1	2
d09	$ 176,600,000	16,000	$ 1.3	$ 3,000	$ 11,000	27,000	10	300	$ 180,000	4	20
d10	$ 4,400,000	12,000	$ 1.0	$ 100	$ 400	1,000	10	200	$ 10,000	10	100
d11	$ 22,700,000	75,000	$ 0.1	$ 100	$ 300	5,000	90	10,000	$ 20,000	10	200
d12	$ 1,800,000	16,000	$ 0.6	$ -	$ 100	400	20	500	$ 2,300	10	30
d13	$ 74,500,000	20,000	$ 0.4	$ 200	$ 4,000	53,000	60	4,000	$ 120,000	10	100
d14	$ 268,500,000	28,000	$ 1.2	$ 8,000	$ 10,000	7,000	1	1	$ 40,000	1	1
d15	$ 129,800,000	89,000	$ 0.1	$ 100	$ 1,000	9,000	20	1,000	$ 130,000	10	200
d16	$ 59,900,000	38,000	$ 0.1	$ 300	$ 2,000	7,000	20	400	$ 80,000	10	100
d17	$ 10,100,000	48,000	$ 0.1	$ 100	$ 200	1,000	20	1,000	$ 10,000	10	100
d18	$ 171,000,000	15,000	$ 6.0	$ 1,000	$ 12,000	119,000	30	1,000	$ 250,000	10	100
d19	$ 899,300,000	460,000	$ 0.1	$ 100	$ 2,000	20,000	50	3,000	$ 350,000	20	300

[3] Recorded amounts can represent a wide range of expenses, incomes, taxes, account balances, etc.

3.2 Estimation Variable

The estimation variable is the error amount (i.e., audit amount, taxable amount) and can be thought of in terms of a tax exposure (i.e., underpayment) or a tax credit (i.e., overpayment). Common audit situation in sales and use tax is where a line item expense, the recorded amount, is either entirely non-taxable (0%) or completely taxable (100%). For this research, this all or nothing scenario was explored with an assumption that 20 percent of the records[4] in the population would be in error (i.e., error rate, discovery rate, population tax rate). All items in the populations have the same probability of being in error with 20 percent of them being randomly assigned as taxable and those assigned as taxable the audit amount equals the recorded amount.

3.3 Upper Dollar Threshold

To account for the unique population attributes of monetary data, an upper dollar threshold is customarily created as the top stratum and reviewed at 100 percent as these "large recorded amounts are of particular auditing concern."[5] (See also Falk & Rotz 2003). This technique has several great qualities: it lessens the non-normality of the data, reduces the required sample size, decreases the variance, does not contribute to the sample error, and limits the projectability of these items to a statistical weight of one further stabilizing the point estimate. There are many names for this top dollar 100 percent review stratum: census, take-all, detail, or certainty stratum. It is a generally accepted sampling technique that is encouraged by practitioners and statisticians and has broad acceptance by regulators and often found in regulatory sampling guidance.

A detail stratum was created for all populations in this research. The threshold is determined judgmentally but quantitatively informed through use of distribution plotting, population coverage, data breaks, and normality statistics. Table 2 provides the before and after skewness and kurtosis statistics for the populations with the top dollar threshold value.

3.4 Sample Design & Sample Sizes

The strata boundaries were constructed, for all non-detail strata, using interval widths and design methods as mentioned above. All methods and populations except population d1 were designed with six strata with the sixth stratum serving as the detail stratum. Population d1 is the same as population d13 but was designed with five strata to link to and build on prior research[6] (Rhyne & Falk 2007) serving as a population and results barometer. Six strata were chosen due to common practices and experience with audit data, existing literature (Cochran 1977; Roberts 1978), and analysis on these populations. Method 8, $cum\sqrt{f}$ 500, due to incompatibility of width size to specific population attributes of m8, m12, and m19 did not run on those populations. In designing with the geometric method, populations m3 - m4, m6 - m7, m8, m11, m14, m16 - m18 needed an extra step and to be analysed with seven to nine strata to accommodate the geometric handling of small values. Once this was done the extra starting strata containing the small dollar amount items were collapsed into one stratum, so all populations ended with six total strata.

---

[4] In sales and use tax audits a 5 percent to 20 percent population error rate is common.

[5] Roberts, D. (1978), "Statistical Auditing," New York, NY: American Institute of Certified Public Accountants, Inc. 1978.

[6] Includes an unpublished ASA Joint Statistical Meeting 2018 Presentation by Rhyne & Pfaffenberger, "Tax Auditing use of Cumulative Square Root of the Frequency Method."

A common practice in audit sampling is to apply constant sample sizes to each non-detail strata, though some states, other regulatory agencies, and consultants use a proportional or optimal allocation method. This constant sample size allows for a standard comparison among the populations and designs and an added applicableness to common audit sampling situations. Frequently seen is the allocation of 100 or 300 to each non-detail stratum. To accommodate the several populations analyzed in this research, a sample size of 80 units were assigned to each non-detail strata for a total sample size of 400 plus the detail stratum. Populations m2, m5, m7, and m8 were exceptions to accommodate how these design methods determined boundaries for those unique population characterises. Populations m2 and m7 were allocated 26 units (for a total of 130 plus detail stratum size) and m5 and m8 were allocated 15 units to all non-detail strata (for a total of 75 units plus detail).

3.5 Simulation Methodology and Random Numbers

One thousand iterations ran for each of the nineteen populations under each of the twelve stratification design methods: 1,000 iterations · 19 populations · 12 methods = 228,000 simulations. The random numbers, sample selections, and iterations ran in SAS. The populations are sorted by the recorded amount. Using SAS' RANUNI function, a random number generator, each iteration for a sample selection was given a seed of 0 (which uses the time of day as the seed). This created a different seed for every iteration applying a random number to each line for each population for every iteration. For each iteration, the population was sorted by strata and the random numbers. The sample was pulled per the designated sample size. Each iteration's sample was evaluated on the taxable amount (i.e., audit finding, the created correlated continuous variable) and projected to the population so that each stratification design method had 1,000 iterations · 12 methods = 12,000 estimates and their associated accuracy measurements.

3.6 Estimation

Among practitioners, auditors, and regulators one of the most commonly used and preferred extrapolation methods, due its simplicity to project sample results, is the Mean per Unit (MPU) estimator (i.e., the direct projection method). This is but one of many generally accepted sampling techniques for extrapolation of audit results to the population which the statistical and audit literature support and can be found in writings like the seminal works of Cochran's *Sampling Techniques* and Roberts' *Statistical Auditing* (which was published by the American Institute of Certified Public Accountants, "AICPA").

The MPU estimator has other great characteristics. "Over all possible samples the mean of the stratified mean estimator equals the total audited amount. Thus, the stratified mean estimator is unbiased."[7] This is a great quality for audit practitioners as estimate amounts often need to be broken out by years, accounts, etc. and this easily reconciles without additional algebraic steps. In most audit and financial documentation reviews, the MPU estimator's "achieved precision" is approximately equal to or larger (i.e., more conservative)[8] than the other generally accepted sampling techniques that meet criteria for use. The MPU estimation technique also does not require observed differences (i.e., the recorded amount less the audit amount) within each stratum (which a lack of observed differences is a frequent audit occurrences).

---

[7] Roberts 1978, p. 101.
[8] Roberts 1978, p. 103.

This is not to say that the MPU estimator is the best audit estimator but possess several robust qualities that serve well for this research and for audit situations in which practitioners have resource constraints. As best practices, it is a generally accepted sampling technique, where feasible, to calculate estimates using multiple techniques and evaluate each on their own merits and comparatively to each other selecting the best in class based on the sample results attributes and other statistical criteria. But it is beyond the scope of this paper to further detail and explore estimators.

For this research, all estimates are produced using the stratified MPU estimator: $\hat{X}_{MS} = \sum(N_i\bar{x}_i)$ and where the estimated standard error is $\hat{\sigma}(\hat{X}_{MS}) = \sqrt{\sum[N_i(N_i - n_i)S_{xi}^2/n_i}$ (Roberts 1978). These formulae's notations are the same as contained in the Internal Revenue Service's Revenue Procedure 2011-42 providing federal "taxpayers with guidance regarding the use and evaluation of statistical samples and sampling estimates." While the notations can vary widely in the literature, the calculations are the same. Roberts' notation of *x* for the audit error amount has often initially confused audit practitioners that are accustomed to seeing *x* represent the recorded amount and *y* the audit error amount. The notation can easily be adjusted as needed if it is plainly documented what it represents.

Relative precision is the margin of error divided by the point estimate with the margin of error being the standard error (i.e., standard deviation) multiplied by the t-value. The 90% confidence limits, built around each estimated audit error amount, are based on the t-value with the degrees of freedom estimated using the Satterthwaite approximation (Cochran 1977).

4. Evaluation of Simulations and Comparison of Methods

To evaluate sample design methods on these specific accounting populations, this analysis looks at estimation and variance accuracy and precision. These attributes are evaluated by analysis on the taxable estimates and those estimates compared to the true taxable amount. Per Cochran, "accuracy refers to the size of deviations from the true mean μ, whereas precision refers to the size of deviations from the mean *m* obtained by repeated application of the sampling procedure."[9] Thus, precision is reliability around producing similar (i.e., consistent) results and accuracy is the closeness of those results to the target (i.e., true value). The assessment considers data levels: the design method level, the population level, and the simulation method level. These statistics, and their derivatives, with the data level views allow for an assessment of the behaviors of the design methods and analysis to conclude on interval width effects for these audit populations.

This paper seeks to disaggregate certain statistics that are often used in audit sampling research. When these statistics are presented at a rolled up, high level, it confounds essential qualities of the estimate, standard deviation, and design. Which confounded qualities should be transparent for use by audit practitioners in understanding best methods for specific circumstances and to prepare audit practitioners for a broader range of statistical behavior when it occurs.

[9] Cochran, W. (1977), "Sampling Techniques, 3rd Edition," New York, NY: John Wiley & Sons, Inc., 1977, p. 16.

Specifically, the following five areas are discussed: (1) mean relative error, (2) relative error frequency, (3) confidence intervals, (4) estimated standard deviation to true taxable amount standard deviation, and (5) mean squared error. The reader should bear in mind during the following discussion that (1) populations 2, 4, and 7 are comparatively large dollar populations by average amount and (2) Method 8 due to incompatibility of interval width to specific population attributes of 8, 9, and 12 did not run on that population and thus had 3,000 less iterations than the other methods. Note than each subsection does not seek to exhaustively discuss the analysis outside of a high-level observation. The ending conclusion will go more in depth on a method by method analysis of observations.

4.1 Mean Relative Error & Relative Error Range Frequency

Table 3 displays that over all iterations by design method and population that on average the mean relative error was essentially 0%. This statistical quality is sought for when choosing statistical techniques and methods. It says that if sampled enough times (and in this study 1,000 times for a population by a method) there is an "accurate" design because the 1,000 estimated taxable amounts average to the true amount. But in practice we do not sample 1,000 times but only once for a project. This balance supports but does not concluded a normality, so that when all errors are weighted over a several random samples, they are all essentially canceled out. Very little else can be ascertained from the table and this aggregated statistic in recommending one method over another or one method best for a type of accounting population. To understand that we need to look at a plot of frequencies or a table, such as Table 4, that list these frequencies with a graphical component.

Table 3: Mean Relative Error

$$\left(Mean\ Relative\ Error = \frac{1}{X}\sum_{i=1}^{X}\frac{\hat{Y}_i - Y}{Y}\right)$$

Data	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	**Population % Mean**
d01	0.1%	-0.1%	0.2%	-0.1%	-0.2%	-0.2%	0.2%	0.0%	0.0%	0.3%	0.1%	0.0%	**0.0%**
d02	-2.5%	-0.4%	0.1%	0.6%	-0.4%	-0.7%	-0.5%	-0.7%	-0.6%	-0.1%	1.6%	1.4%	**-0.2%**
d03	0.0%	-0.2%	-0.3%	0.2%	0.3%	0.0%	-0.3%	-0.2%	0.7%	-0.6%	-0.2%	-0.3%	**-0.1%**
d04	1.1%	-2.8%	-0.5%	-0.4%	0.2%	-0.5%	0.9%	-1.2%	-0.4%	-0.7%	1.1%	1.2%	**-0.2%**
d05	-1.0%	0.3%	0.1%	0.4%	0.7%	-0.5%	-0.8%	0.1%	-1.2%	0.1%	-1.9%	-1.2%	**-0.4%**
d06	0.7%	0.3%	0.3%	-0.4%	0.1%	0.2%	-0.2%	-0.1%	0.1%	0.8%	0.2%	0.2%	**0.2%**
d07	-0.6%	0.6%	0.4%	1.3%	-1.2%	-0.4%	0.1%	0.8%	-0.4%	-1.0%	-0.5%	-1.1%	**-0.2%**
d08	0.4%	-0.3%	0.2%	0.3%	-0.7%	-0.1%	0.1%	*n/a*	-0.3%	0.5%	0.3%	0.1%	**0.0%**
d09	0.5%	-0.7%	0.1%	0.5%	-0.3%	-0.2%	0.1%	0.3%	0.3%	0.3%	0.4%	0.6%	**0.2%**
d10	0.4%	0.3%	-0.1%	-0.1%	0.4%	-0.3%	0.0%	0.1%	-0.3%	-0.2%	0.1%	-0.4%	**0.0%**
d11	1.1%	0.1%	0.3%	0.0%	0.5%	0.3%	-0.2%	0.0%	0.0%	0.7%	0.6%	-1.0%	**0.2%**
d12	0.0%	-0.1%	-0.1%	0.4%	0.1%	-0.2%	-0.2%	*n/a*	0.3%	0.2%	0.1%	0.5%	**0.1%**
d13	-0.1%	0.6%	0.0%	0.6%	0.0%	0.2%	-0.1%	-0.2%	-0.2%	-0.3%	-1.0%	-0.3%	**-0.1%**
d14	0.4%	-0.2%	0.2%	-0.5%	-0.1%	0.1%	0.0%	0.0%	-0.1%	0.8%	0.1%	0.2%	**0.1%**
d15	0.9%	-0.2%	0.6%	-0.6%	0.0%	0.2%	0.2%	0.3%	-0.4%	0.5%	-0.3%	-0.5%	**0.1%**
d16	-0.6%	-0.7%	-0.1%	-0.3%	-0.2%	-0.7%	0.1%	0.0%	-0.6%	0.1%	1.6%	-0.8%	**-0.2%**
d17	-0.9%	-0.3%	0.0%	0.2%	-0.1%	0.4%	0.0%	-0.6%	-0.3%	-0.3%	-0.3%	0.2%	**-0.2%**
d18	0.6%	-0.2%	-0.5%	-0.6%	0.2%	0.0%	-0.1%	0.2%	0.0%	0.6%	0.5%	-0.1%	**0.1%**
d19	1.1%	-0.1%	0.3%	-0.1%	-0.3%	-0.1%	-0.1%	*n/a*	-0.3%	0.8%	1.1%	0.0%	**0.2%**
Method % Mean	**0.1%**	**-0.2%**	**0.1%**	**0.1%**	**-0.1%**	**-0.1%**	**0.0%**	**-0.1%**	**-0.2%**	**0.1%**	**0.2%**	**-0.1%**	**0.0%**

Table 4a breaks down Table 3 to the iteration level within method showing the frequencies of the relative errors within 25% intervals. Here data shape is transparent and assist in understanding the errors of over and under estimating the true taxable amount. The values of the frequency of the relative error show normal distribution tendencies, peaking strongly around the true taxable amount with a tight -25% to 25% error range. There visually appears, and in the associated count summation, to be roughly 50% of estimates on both sides of 0% (or the true taxable amount) but the length and the tails are not symmetrical. The table shows that even though both sides are balanced in the taxable error weight the data does not all exhibit proportional shape with overestimating the true value having a long downward tail (i.e., skewed to the right for a graph). The extreme outliers occur infrequently but can overestimate the true error rate by up to 425%, an extremely exaggerated error level (a distinct challenge of dealing with accounting data).

Table 4a: Relative Error Frequency

Range	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	**Population Total**	**Population % Mean**
(-75%) - (-50%)	135	150	1	92	23	21	4	3	3	15	261	135	843	0.4%
(-50%) - (-25%)	2,011	1,381	860	1,060	726	621	571	536	128	1,086	2,544	1,797	13,321	5.9%
(-25%) - 0%	8,151	8,396	8,973	8,556	8,813	8,807	8,845	7,446	9,436	8,550	7,686	8,336	101,995	45.3%
0% - 25%	6,459	7,625	8,071	8,132	8,620	8,859	8,862	7,304	9,182	7,823	5,942	6,609	93,488	41.6%
25% - 50%	1,592	1,039	828	784	593	520	559	595	248	1,337	1,670	1,544	11,309	5.0%
50% - 75%	377	217	186	209	144	116	105	89	3	165	457	340	2,408	1.1%
75% - 100%	144	92	57	84	53	37	38	19		19	207	116	866	0.4%
100% - 125%	69	48	15	47	15	13	14	8		5	83	59	376	0.2%
125% - 150%	24	26	6	15	6	2	1				45	20	145	0.1%
150% - 175%	9	13	2	6	4	3					32	20	89	0.0%
175% - 200%	10	5	1	7	1	1	1				27	7	60	0.0%
200% - 225%	5	3		6	2						24	8	48	0.0%
225% - 250%	5	3		1							9	3	21	0.0%
250% - 275%	7	1		1							6	4	19	0.0%
275% - 300%	1	1									1		3	0.0%
300% - 325%											3	2	5	0.0%
325% - 350%											2		2	0.0%
350% - 375%														0.0%
375% - 400%	1										1		2	0.0%
400% - 425%														0.0%
Method Total	**19,000**	**19,000**	**19,000**	**19,000**	**19,000**	**19,000**	**19,000**	**16,000**	**19,000**	**19,000**	**19,000**	**19,000**	**225,000**	**100.0%**

True Taxable Amount

What at first, from Table 3, appears no one estimator projecting more accurate estimates, or no distinguish effect for the interval widths, Table 4 clearly exposes an interval width effect and one that runs counter to existing guidance of having narrow intervals (i.e., many classes). (The authors have initially used a table and not a graph in the belief that graphs and plots are essential but overused in data analysis at the expense of tables. Tables can often preserve more information, and more cleanly, than graphs and exhibit it in ways that uniquely and unexpectedly emphasis the data's shape graphically.) While serving as a barometer, the Equal $ Method is the most accurate design method for the specific situations. It has a tight error range, a strong peak, and no comparatively extreme projection errors. No other tested method approaches the Equal $ Method in exhibiting this level of accurateness or precision in estimating the true taxable amount.

This is not the only design quality that should be explored. Before going on to the other areas of assessment this analysis removes the comparatively extreme large dollar audit populations 2, 4, and 7 from this frequency table as seen in Table 4b. (See accounting population Table 2 for further details on these specific populations.) Without those large dollar populations, some of the other design methods and interval widths look to perform as well as the Equal $ Method. Here the advantages of the Equal $ Method are vastly lessened though it continues to exhibit the strongest normal tendencies of all the methods. The effect of interval widths continues to be present in excluding large dollar populations.

Table 4b: Relative Error Frequency (Excluding the large dollar populations)

Range	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	**Population Total**	**Population % Mean**
-100% -75%														0.0%
-75% -50%	20	17	1	8					3	12	132	17	210	0.1%
-50% -25%	1,188	596	297	337	141	126	123	167	113	694	1,713	966	6,461	3.4%
-25% 0%	7,211	7,464	7,770	7,575	7,657	7,600	7,613	6,139	7,778	7,283	6,776	7,401	88,267	46.7%
0% 25%	5,905	7,069	7,349	7,544	7,950	8,084	8,048	6,416	7,936	6,922	5,384	6,196	84,803	44.9%
25% 50%	1,358	785	548	503	248	185	210	268	167	1,002	1,402	1,227	7,903	4.2%
50% 75%	248	63	34	31	4	5	6	10	3	79	343	164	990	0.5%
75% 100%	47	6	1	2						7	148	24	235	0.1%
100% 125%	22									1	54	5	82	0.0%
125% 150%											24		24	0.0%
150% 175%	1										9		10	0.0%
175% 200%											6		6	0.0%
200% 225%											5		5	0.0%
225% 250%											1		1	0.0%
250% 275%											3		3	0.0%
Method Total	**16,000**	**16,000**	**16,000**	**16,000**	**16,000**	**16,000**	**16,000**	**13,000**	**16,000**	**16,000**	**16,000**	**16,000**	**189,000**	**100.0%**

True Taxable Amount

4.2 Confidence Interval Coverage

The 90% confidence intervals are built to review the variance and normality assumptions. Table 5 show the percentage of times, the true taxable amount is contained within the estimated taxable amount interval. The average coverage among methods of the true taxable amount ranged from 79% to 88% and on average 85%. The effect of interval widths can be seen with the wider intervals being comparative to the equal $ amount method. Though once again the equal $ method performs most consistently among all populations demonstrating its robustness exceptionally well on the extraordinarily large dollar populations where the other methods and specific interval widths fell short. Without the large dollar populations, the average coverage is 87%. Given some of the small sample sizes based on population attributes, the potential for some designs being less effective than others, and the skewness of accounting data in general, there was an expectation that the coverage rate would fall below but remain close to 90%.

Table 5: Confidence Interval Coverage of the True Taxable Amount
(Extraordinarily large dollar populations in red)

Data	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	**Population % Mean**	Stratum non-detail **Sample Size**
d01	81%	83%	87%	89%	88%	90%	90%	92%	87%	84%	78%	81%	**86%**	80
d02	56%	58%	59%	63%	69%	71%	74%	76%	80%	75%	50%	64%	**66%**	26
d03	87%	85%	89%	86%	90%	90%	89%	90%	89%	86%	85%	87%	**88%**	80
d04	67%	68%	79%	68%	73%	78%	78%	76%	87%	79%	70%	69%	**74%**	80
d05	86%	85%	91%	89%	90%	91%	91%	91%	89%	89%	86%	88%	**89%**	15
d06	90%	88%	89%	88%	91%	90%	91%	90%	90%	89%	89%	90%	**90%**	80
d07	66%	66%	73%	67%	69%	69%	70%	74%	87%	75%	72%	59%	**71%**	26
d08	86%	86%	86%	88%	88%	87%	90%	n/a	89%	86%	89%	87%	**87%**	15
d09	87%	87%	89%	91%	88%	92%	89%	90%	90%	89%	88%	90%	**89%**	80
d10	88%	90%	89%	89%	90%	92%	90%	88%	90%	89%	87%	86%	**89%**	80
d11	76%	86%	88%	88%	91%	90%	90%	88%	89%	88%	72%	77%	**85%**	80
d12	88%	90%	91%	91%	88%	89%	88%	n/a	90%	89%	87%	90%	**89%**	80
d13	76%	85%	85%	88%	88%	90%	90%	89%	88%	88%	73%	77%	**85%**	80
d14	89%	89%	88%	88%	90%	90%	90%	90%	89%	89%	91%	89%	**89%**	80
d15	83%	87%	86%	87%	90%	91%	92%	90%	88%	87%	74%	81%	**86%**	80
d16	79%	81%	86%	88%	89%	90%	90%	90%	89%	88%	78%	79%	**86%**	80
d17	83%	89%	89%	91%	89%	92%	89%	85%	89%	88%	80%	85%	**87%**	80
d18	83%	83%	88%	86%	90%	89%	91%	88%	90%	89%	81%	84%	**87%**	80
d19	78%	86%	86%	87%	87%	90%	90%	n/a	88%	85%	68%	81%	**84%**	80
Method % Mean	**80%**	**83%**	**85%**	**85%**	**86%**	**87%**	**87%**	**87%**	**88%**	**86%**	**79%**	**81%**	**85%**	

4.3 Estimated Standard Deviation and True Standard Deviation

To assess the design methods' standard deviation behavior, the percent difference between sample standard deviation and the true standard deviation is listed in below Table 6 and calculated as $(Sample\ Error - True\ Standard\ Deviation) \div True\ Standard\ Deviation.$ (Note that when this activity is done for the variance differences the percent errors difference balances out to essentially 0%.) When viewing the Table 6, recall that m08 has 3,000 less iterations.

The table shows a normal tendency to the frequencies with the interval widths showing tighter tails and higher peaks as the interval width widens. The equal $ method shows a unique characteristic compared to the other methods. It is far less inclined to underestimate the true standard deviation and one of the highest peaks at the -25% to 25% range and among the tightest distributions. It does have a long but thin tail of overestimating the variance. The Geometric method gave the least occurrences of and the shortest tail in overestimating errors though its central peak is not as high as other methods.

Table 6: Frequency of Difference between Sample and Population Standard Deviation

Range	(cum. √f): 0.01 m01	(cum. √f): 1 m02	(cum. √f): median/100 m03	(cum. √f): 5 m04	(cum. √f): 50 m05	(cum. √f): 100 m06	(cum. √f): 250 m07	(cum. √f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	Total
(-100%) - (-75%)	829	746	420	585	323	250	169	95		53	1188	805	5463
(-75%) - (-50%)	2403	1164	794	878	851	765	712	685	396	767	2766	2164	14345
(-50%) - (-25%)	3522	2743	2454	1916	747	737	945	1198	1376	2271	3448	3327	24684
(-25%) - 0%	4855	6189	6821	7048	7973	7956	7938	6187	8639	7113	4545	5303	80567
0% - 25%	4307	5789	6547	6812	8263	8598	8313	6753	7589	6790	4120	4458	78339
25% - 50%	1741	1686	1365	1304	380	276	479	744	790	1447	1502	1667	13381
50% - 75%	680	390	367	155	226	254	248	222	134	404	673	716	4469
75% - 100%	315	111	97	90	129	74	124	91	20	123	337	226	1737
100% - 125%	142	49	48	77	49	54	51	18	20	23	122	159	812
125% - 150%	93	50	32	42	38	21	17	4	14	7	106	90	514
150% - 175%	38	19	46	58	14	10	2	3	3	2	54	41	290
175% - 200%	32	31	4	22	4	3	1		10		50	18	175
200% - 225%	21	23		7	2	2	1		4		11	7	78
225% - 250%	9	6	3	2	1				4		19	10	54
250% - 275%	6	3	2	1					1		19	6	38
275% - 300%	6			3							19	3	31
300% - 325%		1									1		2
325% - 350%											18		18
350% - 375%											2		2
375% - 400%	1												1
Min	-98%	-99%	-97%	-98%	-97%	-96%	-97%	-94%	-71%	-96%	-100%	-100%	-100%
Mean	**-9.8%**	**-6.8%**	**-4.7%**	**-5.0%**	**-3.4%**	**-3.1%**	**-2.7%**	**-3.1%**	**-2.7%**	**-3.1%**	**-11.9%**	**-9.2%**	**-5.5%**
Max	396%	318%	267%	287%	232%	215%	201%	170%	268%	169%	353%	299%	396%
Total # of Iterations	19,000	19,000	19,000	19,000	19,000	19,000	19,000	16,000	19,000	19,000	19,000	19,000	225,000
Total # above 100%	348	182	135	212	108	90	72	25	56	32	421	334	2,015

True Standard Deviation

4.4 Relative Precision

Per Roberts, precision is a "measure of closeness between a sample estimate and the corresponding unknown population characteristic" (Roberts 1978). It is a widely used measurement in audit sampling and when certain thresholds of precision (e.g., 10%) are not met the taxpayer often must us the lower bound of the confidence interval (Falk, Petska, & Rhyne 2008). For the degrees of freedom (df), the Satterthwaite approximation was used (Cochran 1977). Here again the incrementally increasing improved effect is seen for the widening interval widths. The Equal $ Method stands out as the best preforming method on these methods with a mean relative precision of 15.3%. The Geometric method comparatively performs poorly.

Table 7: Mean Relative Precision by Method & Population

($Relative\ Precision = \frac{t(df)\sqrt{v(\hat{Y})}}{(\hat{Y})}$ where Y is the taxable amount)

Data	(cum. √f): 0.01 m01	(cum. √f): 1 m02	(cum. √f): median/100 m03	(cum. √f): 5 m04	(cum. √f): 50 m05	(cum. √f): 100 m06	(cum. √f): 250 m07	(cum. √f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	**Population % Mean**
d01	15.9%	11.2%	10.1%	9.1%	7.1%	6.8%	6.7%	6.7%	7.5%	13.2%	16.7%	15.4%	**10.5%**
d02	31.3%	33.3%	31.9%	32.3%	27.9%	25.9%	23.8%	22.0%	12.8%	25.2%	35.1%	33.0%	**27.9%**
d03	32.5%	30.1%	20.8%	27.2%	20.0%	18.0%	15.8%	14.3%	12.9%	30.1%	35.2%	31.6%	**24.0%**
d04	59.2%	58.0%	37.4%	55.3%	48.1%	45.3%	41.9%	38.1%	19.1%	39.4%	60.3%	59.0%	**46.8%**
d05	43.4%	42.7%	36.0%	39.7%	32.7%	31.2%	30.5%	31.4%	30.4%	30.1%	43.4%	43.2%	**36.2%**
d06	23.2%	22.0%	17.1%	19.4%	15.7%	15.3%	14.9%	14.9%	14.2%	24.8%	24.2%	22.7%	**19.0%**
d07	42.3%	42.6%	37.5%	42.4%	40.4%	39.6%	38.4%	37.7%	19.8%	36.6%	36.5%	44.9%	**38.2%**
d08	22.7%	21.6%	20.1%	18.6%	15.4%	14.3%	13.7%	*n/a*	13.8%	27.9%	21.0%	23.0%	**19.3%**
d09	31.0%	31.0%	21.9%	27.7%	20.9%	19.5%	18.1%	17.4%	16.9%	33.9%	35.4%	28.4%	**25.2%**
d10	25.0%	15.8%	15.2%	13.6%	12.6%	12.7%	13.2%	18.1%	12.6%	18.8%	25.4%	24.3%	**17.3%**
d11	32.3%	18.9%	20.4%	15.8%	13.9%	13.8%	14.8%	19.1%	13.8%	21.3%	38.4%	28.1%	**20.9%**
d12	26.6%	16.0%	17.5%	15.0%	14.8%	16.2%	24.1%	*n/a*	14.8%	21.0%	30.9%	24.3%	**20.1%**
d13	23.7%	17.4%	15.1%	13.7%	9.5%	8.8%	8.1%	7.8%	8.5%	14.5%	24.2%	22.5%	**14.5%**
d14	20.9%	19.8%	18.2%	18.6%	18.2%	18.2%	18.1%	18.1%	17.4%	32.4%	21.0%	21.0%	**20.1%**
d15	40.9%	26.5%	26.8%	20.8%	17.0%	16.4%	16.0%	15.8%	17.5%	31.8%	47.9%	37.0%	**26.2%**
d16	33.1%	25.0%	21.7%	20.6%	16.1%	15.3%	14.6%	14.9%	14.9%	22.7%	38.2%	31.3%	**22.4%**
d17	29.8%	18.6%	19.7%	16.5%	15.3%	15.3%	18.2%	26.4%	15.6%	25.5%	35.6%	27.1%	**22.0%**
d18	25.6%	23.1%	17.0%	19.4%	14.1%	12.6%	11.2%	10.7%	10.4%	15.0%	29.1%	24.1%	**17.7%**
d19	42.4%	28.2%	26.8%	22.5%	17.3%	16.5%	16.1%	*n/a*	17.7%	33.9%	56.6%	35.5%	**28.5%**
Method % Mean	**31.7%**	**26.4%**	**22.7%**	**23.6%**	**19.8%**	**19.0%**	**18.8%**	**19.6%**	**15.3%**	**26.2%**	**34.5%**	**30.3%**	**24.1%**
- without 2, 4,7	29.3%	23.0%	20.3%	19.9%	16.3%	15.7%	15.9%	16.6%	14.9%	24.8%	32.7%	27.5%	21.5%
Total # of Iterations	19,000	19,000	19,000	19,000	19,000	19,000	19,000	16,000	19,000	19,000	19,000	19,000	225,000

As above with the other analysis, the following Table 8a and Table 8b will disaggregate this statistic to the iteration level by method and view it with and without the large dollar populations where the same effects and conclusions can be observed as already identified.

Table 8a: Relative Precision Frequency Ranges

Range	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	Population Total	Population % Mean
0% - 10.0%	364	478	555	821	1,547	1,962	2,077	2,048	2,437	184	485	383	13,341	6%
10% - 20.0%	2,117	6,337	8,649	9,097	12,315	12,847	12,174	8,773	14,129	4,454	1,953	2,489	95,334	42%
20% - 30.0%	8,085	6,959	6,376	5,552	2,373	1,662	2,326	2,518	1,750	7,971	6,361	8,697	60,630	27%
30% - 40.0%	4,286	2,842	1,985	1,436	1,336	1,213	1,352	1,744	656	4,950	4,555	4,050	30,405	14%
40% - 50.0%	2,034	1,003	804	874	477	535	456	472	28	1,116	2,963	1,746	12,508	6%
50% - 60.0%	994	454	436	464	567	503	446	383		258	1,208	671	6,384	3%
60% - 70.0%	630	504	82	399	260	179	121	50		56	570	315	3,166	1%
70% - 80.0%	242	185	53	183	62	59	44	12		11	346	260	1,457	1%
80% - 90.0%	101	93	46	82	35	30	4				229	225	845	0%
90% - 100.0%	71	74	14	37	18	8					138	88	448	0%
100% - 110.0%	38	40		27	7	2					86	39	239	0%
110% - 120.0%	18	14		22	3						51	23	131	0%
120% - 130.0%	15	12		6							37	12	82	0%
130% - 140.0%	5	5									18	2	30	0%
Total # of Iterations	19,000	19,000	19,000	19,000	19,000	19,000	19,000	16,000	19,000	19,000	19,000	19,000	225,000	100%
Total # above 40%	4,148	2,384	1,435	2,094	1,429	1,316	1,071	917	28	1,441	5,646	3,381	25,290	

Table 8b: Relative Precision Frequency Ranges
(Excluding the large dollar populations)

Range	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	Population Total	Population % Mean
0% - 10.0%	212	319	429	691	1,460	1,886	2,001	2,000	1,908	161	239	224	11,530	6%
10% - 20.0%	1,764	5,991	8,186	8,726	11,891	12,393	11,675	8,173	12,608	3,962	1,588	2,137	89,094	47%
20% - 30.0%	7,534	6,432	5,802	5,057	1,831	1,046	1,677	1,924	892	7,362	6,019	8,247	53,823	28%
30% - 40.0%	3,951	2,505	1,342	1,056	775	667	639	888	579	4,099	4,213	3,564	24,278	13%
40% - 50.0%	1,649	629	234	423	43	8	8	15	13	394	2,317	1,395	7,128	4%
50% - 60.0%	644	108	7	45						22	808	360	1,994	1%
60% - 70.0%	170	13		2							360	63	608	0%
70% - 80.0%	64	2									186	9	261	0%
80% - 90.0%	10	1									133	1	145	0%
90% - 100.0%	2										72		74	0%
100% - 110.0%											39		39	0%
110% - 120.0%											20		20	0%
120% - 130.0%											5		5	0%
130% - 140.0%											1		1	0%
Total # of Iterations	16,000	16,000	16,000	16,000	16,000	16,000	16,000	13,000	16,000	16,000	16,000	16,000	189,000	100%
Total # above 40%	2,539	753	241	470	43	8	8	15	13	416	3,941	1,828	10,275	

4.5 Mean Squared Error

As the Mean Squared Error statistics is not a relative measurement and will not be averaged among the population but only across the design methods. The Mean Squared Error equals $(variance\ of\ taxable\ estiamte) + (bias)^2$ where bias is the taxable estimate less the true taxable amount (Cochran 1977). The same patterns as seen above hold in this Table 9.

Table 9: Mean Relative Precision by Method & Population

Data	(cum. √ f): 0.01 m01	(cum. √ f): 1 m02	(cum. √ f): median/100 m03	(cum. √ f): 5 m04	(cum. √ f): 50 m05	(cum. √ f): 100 m06	(cum. √ f): 250 m07	(cum. √ f): 500 m08	Equal $ Amount m09	Geometric m10	Equal Nh m11	Equal Nh per unique $ m12	Population Mean
d01	6.0E+12	2.6E+12	2.0E+12	1.5E+12	9.4E+11	8.7E+11	8.6E+11	8.0E+11	1.1E+12	3.8E+12	7.3E+12	5.3E+12	5.3E+12
d02	6.0E+16	6.4E+16	5.6E+16	4.8E+16	2.4E+16	1.9E+16	1.5E+16	1.1E+16	4.6E+15	1.6E+16	1.6E+17	5.4E+16	5.4E+16
d03	9.4E+13	8.3E+13	3.7E+13	6.7E+13	3.4E+13	2.7E+13	2.1E+13	1.7E+13	1.5E+13	8.2E+13	1.2E+14	8.9E+13	8.9E+13
d04	1.9E+20	1.6E+20	4.0E+19	1.5E+20	8.6E+19	6.6E+19	5.6E+19	4.2E+19	8.2E+18	4.2E+19	1.9E+20	1.8E+20	1.8E+20
d05	8.4E+09	8.7E+09	5.7E+09	7.2E+09	4.9E+09	4.3E+09	4.3E+09	4.5E+09	4.3E+09	4.2E+09	8.7E+09	8.3E+09	8.3E+09
d06	2.1E+11	1.9E+11	1.2E+11	1.5E+11	9.7E+10	9.2E+10	8.6E+10	8.8E+10	7.9E+10	2.5E+11	2.3E+11	2.0E+11	2.0E+11
d07	4.9E+17	5.3E+17	3.4E+17	5.2E+17	4.2E+17	4.0E+17	3.8E+17	3.2E+17	6.8E+16	2.7E+17	3.0E+17	8.2E+17	8.2E+17
d08	3.7E+08	3.3E+08	3.0E+08	2.5E+08	1.7E+08	1.5E+08	1.3E+08	*n/a*	1.4E+08	5.9E+08	3.0E+08	3.8E+08	3.8E+08
d09	8.8E+13	8.6E+13	4.2E+13	6.7E+13	3.8E+13	3.2E+13	2.8E+13	2.7E+13	2.4E+13	1.0E+14	1.1E+14	7.2E+13	7.2E+13
d10	3.7E+10	1.4E+10	1.2E+10	1.0E+10	8.7E+09	8.3E+09	9.5E+09	1.8E+10	8.4E+09	1.9E+10	3.9E+10	3.4E+10	3.4E+10
d11	2.3E+12	5.2E+11	6.0E+11	3.4E+11	2.6E+11	2.6E+11	2.9E+11	5.1E+11	2.6E+11	6.5E+11	4.0E+12	1.5E+12	1.5E+12
d12	6.6E+09	2.3E+09	2.7E+09	2.0E+09	2.0E+09	2.3E+09	5.4E+09	*n/a*	1.9E+09	4.0E+09	9.2E+09	5.5E+09	5.5E+09
d13	1.5E+13	6.4E+12	4.8E+12	3.7E+12	1.7E+12	1.4E+12	1.2E+12	1.1E+12	1.4E+12	4.0E+12	1.7E+13	1.3E+13	1.3E+13
d14	8.7E+13	8.0E+13	6.8E+13	7.1E+13	6.6E+13	6.7E+13	6.5E+13	6.6E+13	6.0E+13	2.1E+14	8.5E+13	8.9E+13	8.9E+13
d15	1.3E+14	4.2E+13	4.7E+13	2.6E+13	1.6E+13	1.5E+13	1.4E+13	1.4E+13	1.8E+13	6.4E+13	2.2E+14	1.0E+14	1.0E+14
d16	1.6E+13	8.2E+12	5.6E+12	4.9E+12	3.0E+12	2.5E+12	2.3E+12	2.4E+12	2.5E+12	5.8E+12	2.5E+13	1.3E+13	1.3E+13
d17	3.1E+11	1.0E+11	1.1E+11	8.1E+10	7.0E+10	6.6E+10	1.0E+11	2.2E+11	7.3E+10	1.9E+11	5.0E+11	2.5E+11	2.5E+11
d18	6.3E+13	4.9E+13	2.4E+13	3.2E+13	1.6E+13	1.3E+13	9.8E+12	9.9E+12	8.7E+12	1.9E+13	9.0E+13	5.3E+13	5.3E+13
d19	6.7E+15	2.2E+15	2.0E+15	1.3E+15	7.8E+14	6.9E+14	6.2E+14	*n/a*	8.3E+14	3.3E+15	1.9E+16	4.0E+15	4.0E+15

5. Conclusion

The research supports the conclusion that interval width (i.e., class width) has a meaningful effect on the $cum\sqrt{f}$ method and the associated representativeness of the sample and the accuracy and precision of the estimate. In much of the existing literature on sample designs, in which stratification methods have been compared to the $cum\sqrt{f}$ method, there must be a consideration that the interval widths in those analysis were potentially far from appropriate for those populations and further consideration should be given where no explanation was provided for the chosen width(s).

Cochran wrote, which he applied to a different issue, but one stated here: "Only a few examples from actual data are available in the literature. To supplement them, a simple theoretical approach is used."[10] This research does the opposite. In discussing interval widths, this research does not seek a theoretical approach, but actual data driven analysis with more populations than is characteristically seen in applicable research. But nineteen populations are still a small number of (non-random) populations and there is risk they don't represent the broader universe of accounting data.

With these populations, the larger interval widths resulted in better statistical measurements which runs counter to common guidance in the literature of narrow interval widths (or "many" classes). The similarities in performance of intervals $0.01 and $1 to the corresponding Equal N_h and Equal N_h per unique $ stood out in the analysis and supports the conclusion that narrow widths is not, in general, an effective strategy for highly skewed, large dollar populations.

But this research is not for concluding on best design methods or optimal interval widths. These interval widths were not chosen to maximize the best statistical qualities but to explore narrow, moderate, and wide interval widths and to conclude on their general effects with the $cum\sqrt{f}$ method. The other methods analyzed in this research are used as barometers for the $cum\sqrt{f}$ method and those methods performance compared to the $cum\sqrt{f}$ method does not allow for any generalization on them being better or worse methods outside of conclusions on the specific interval widths used in respect to the populations considered in this study.

It is our opinion the $cum\sqrt{f}$ method is a highly effective sample design method for highly skewed accounting data when used dynamically toward an understanding of a population's unique attributes and consciously to create desirable statistical qualities such as equalizing some product of S_h among strata (Cochran 1961; Cochran 1977). Of interest is how a simple interval width formula, median/100, performed in relation to the other interval widths and encourages looking into a formula that considers more suitable population attributes.

The Geometric method often exhibited distributions, in both estimate error and standard deviation error, that were not extremely skewed with or without the large dollar populations. It showed a tighter but flatter (i.e., not as high of a peak around the center) distribution. Comparatively, it had tendencies to underestimate the true taxable amount with the larger populations and overestimate with the smaller populations. It was among the poorest performers in relative precision.

---

[10] Cochran, W. (1977), "Sampling Techniques, 3rd Edition," New York, NY: John Wiley & Sons, Inc., 1977, p. 132.

The Equal $ Method showed a superior performance among the large dollar populations and among the best performance in most circumstances. It had higher and tighter relative error frequency distributions. What is very interesting is in Table 6 where it clearly shows its ability to control for underestimating the true standard deviation though it has a long but extremely thin tail for overestimating standard deviation. While it is skewed to the right, it still had less occurrences than all other methods above 50% difference in overestimating standard deviation. Its great statistical performance continued with relative precision. This is not to recommend the Equal $ Method above the $cum\sqrt{f}$ method as again this research was not centered on identifying optional interval widths for the $cum\sqrt{f}$ method. Table 9 and Table 10 provide some additional high level summaries of the analysis.

Table 9: Mean Relative Precision and Mean Relative Error Summary

								Without populations 2, 4, & 7					
		Mean Relative Error			Mean Relative Precision			Mean Relative Error			Mean Relative Precision		
ID	Description	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
m01	(cum. √ f): 0.01	-2.5%	1.1%	0.1%	15.9%	59.2%	31.7%	-1.0%	1.1%	0.2%	15.9%	43.4%	29.3%
m02	(cum. √ f): 1	-2.8%	0.6%	-0.2%	11.2%	58.0%	26.4%	-0.7%	0.6%	-0.1%	11.2%	42.7%	23.0%
m03	(cum. √ f): median/100	-0.5%	0.6%	0.1%	10.1%	37.5%	22.7%	-0.5%	0.6%	0.1%	10.1%	36.0%	20.3%
m04	(cum. √ f): 5	-0.6%	1.3%	0.1%	9.1%	55.3%	23.6%	-0.6%	0.6%	0.0%	9.1%	39.7%	19.9%
m05	(cum. √ f): 50	-1.2%	0.7%	-0.1%	7.1%	48.1%	19.8%	-0.7%	0.7%	0.0%	7.1%	32.7%	16.3%
m06	(cum. √ f): 100	-0.7%	0.4%	-0.1%	6.8%	45.3%	19.0%	-0.7%	0.4%	-0.1%	6.8%	31.2%	15.7%
m07	(cum. √ f): 250	-0.8%	0.9%	0.0%	6.7%	41.9%	18.8%	-0.8%	0.2%	-0.1%	6.7%	30.5%	15.9%
m08	(cum. √ f): 500	-1.2%	0.8%	-0.1%	6.7%	38.1%	19.6%	-0.6%	0.3%	0.0%	6.7%	31.4%	16.6%
m09	Equal $ Amount	-1.2%	0.7%	-0.2%	7.5%	30.4%	15.3%	-1.2%	0.7%	-0.1%	7.5%	30.4%	14.9%
m10	Geometric	-1.0%	0.8%	0.1%	13.2%	39.4%	26.2%	-0.6%	0.8%	0.3%	13.2%	33.9%	24.8%
m11	Equal Nh	-1.9%	1.6%	0.2%	16.7%	60.3%	34.5%	-1.9%	1.6%	0.1%	16.7%	56.6%	32.7%
m12	Equal Nh per unique $	-1.2%	1.4%	-0.1%	15.4%	59.0%	30.3%	-1.2%	0.6%	-0.2%	15.4%	43.2%	27.5%

Table 10: Mean Relative Error Range Coverage and Best Relative Precision by Method

		Relative Error			Mean Relative Precision			
ID	Description	(-5%) - 5%	(-15%) - 15%	Mean Relative Precision	1st best	2nd best	3rd best	% of Top 3
m01	(cum. √ f): 0.01	20%	54%	31.7%	0	0	0	0%
m02	(cum. √ f): 1	26%	65%	26.4%	0	0	0	0%
m03	(cum. √ f): median/100	30%	72%	22.7%	0	1	0	5%
m04	(cum. √ f): 5	31%	71%	23.6%	0	0	1	5%
m05	(cum. √ f): 50	37%	78%	19.8%	1	2	2	26%
m06	(cum. √ f): 100	39%	81%	19.0%	1	2	4	37%
m07	(cum. √ f): 250	40%	81%	18.8%	3	3	7	68%
m08	(cum. √ f): 500	38%	78%	19.6%	3	7	1	58%
m09	Equal $ Amount	44%	88%	15.3%	10	3	3	84%
m10	Geometric	26%	66%	26.2%	1	0	1	11%
m11	Equal Nh	18%	49%	34.5%	0	1	0	5%
m12	Equal Nh per unique $	21%	56%	30.3%	0	0	0	0%

Acknowledgements

The authors would like to thank the following people and organizations for their contributions: Kimberly Rhyne; Eric Falk who co-authored with Zac Rhyne the 2007 Joint Statistical Meeting (JSM) paper which initial research this paper has continued to build on; Ryan LLC, Team Sampling members: Jason McGlamery and Michael Putin; the statistical auditing interest group of ASA for its great collaboration, sponsoring of audit sampling sessions, and members Wendy Rotz, Ryan Petska, Ed Cohen, and Mary Batcher.

Additionally, there is a singular recognition of the co-author Roger Pfaffenberger who passed away prior to the completion of this work and in whom I have the deepest respect for his kind and tremendous influence. Our work on this research began with our JSM

2018 presentation which findings we determined to be interesting and exceptionally relevant to further explore forming the foundation of this more extensive JSM 2019 research.

References

Cochran, W.G. (1961), “Comparison of methods for determining stratum boundaries,” Bulletin of the International Statistical Institute, 32, 2, 345-358.

Cochran, W. G. (1977), “Sampling Techniques, 3rd Edition,” New York, NY: John Wiley & Sons, Inc., 1977.

Dalenius, T. & Hodges, J. (1959), “Minimum variance stratification, “Journal of the American Statistical Association, 88-101.

Falk, E., & Rotz, W. (2003), “Stratified sampling for sales and use tax highly skewed data-determination of the certainty stratum cut-off amount,” Proceedings of the Section on Statistical Computing, American Statistical Association, 66-72.

Falk, E., Petska, R., & Rhyne, Z. (2008), “How to help businesses make regulatory sampling decision,” 2008 Proceedings of the American Statistical Association.

Gunning, P. & Horgan J. (2004), “A new algorithm for the construction of stratum boundaries in skewed populations,” Survey Methodology 2, p.30.

Gunning, P., Horgan, J., & Yancey, W (2004), “Geometric stratification of accounting data,” Contaduría y Administración, 214.

Hedlin, D. (2000), “A procedure for stratification by an extended Ekman rule,” Journal of Official Statistics, 16, 15-29.

Horgan, J. (2010), “Choosing the stratification boundaries: The elusive optima,” Istanbul University Journal of the School of Business Administration, 195 – 204.

Rhyne, Z. & Falk, E. (2007), “Handling negatives in statistical samples of business records,” 2007 Proceedings of the American Statistical Association.

Roberts, D. (1978), “Statistical Auditing,” New York, NY: American Institute of Certified Public Accountants, Inc. 1978.