Appendix L

Alternate Stratification Methods

Contents:

Equal Invoice Amount

Geometric Ratio

In stratifying the population, the most common method found in sales and use tax sampling is shown in Appendix G as optimal sampling (Cumulative Square Root of the Frequency). However, there are other methods available in establishing the strata boundaries. Most of these deal with some amount, usually the invoice or other book amount that is available to the auditor.

Equal Invoice Amount[1]

This method is simple, in that it tries to allocate the same total invoice value within each stratum. The population is sorted from smallest to largest invoice value. The total value is divided by the number of strata to determine the first break value. The invoice value, accumulated from the lowest value that matches this value, is the last invoice in the first stratum. The next invoice belongs to the next stratum. This process is continued throughout the sampling frame. An example of this methodology is as follows:

[1] This method is discussed in Roberts' *Statistical Auditing* at page 97. He refers to it as "equal recorded amount".

Equal Invoice Amount

Population Total	31,870
Population Count	40
Break Value	10,623.33

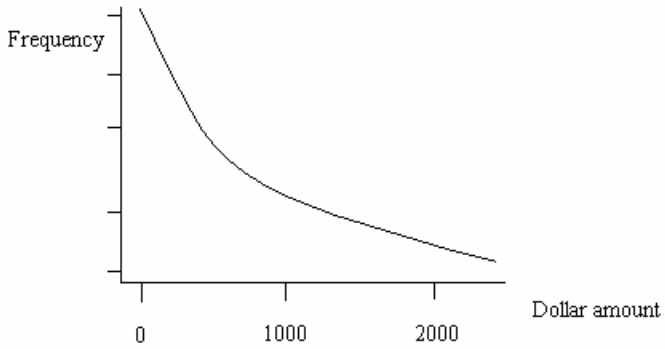
Stratum Element	Invoice Value	Stratum	Cumulative Total
1	100	1	100
2	101	1	201
3	102	1	303
4	103	1	406
5	104	1	510
6	300	1	810
7	310	1	1,120
8	320	1	1,440
9	330	1	1,770
10	340	1	2,110
11	350	1	2,460
12	420	1	2,880
13	750	1	3,630
14	760	1	4,390
15	770	1	5,160
16	780	1	5,940
17	790	1	6,730
18	800	1	7,530
19	810	1	8,340
20	820	1	9,160
21	830	1	9,990
22	840	1	10,830
1	850	2	850
2	910	2	1,760
3	1,100	2	2,860
4	1,110	2	3,970
5	1,120	2	5,090
6	1,130	2	6,220
7	1,160	2	7,380
8	1,170	2	8,550
9	1,180	2	9,730
10	1,190	2	10,920
1	1,200	3	1,200
2	1,210	3	2,410
3	1,220	3	3,630
4	1,230	3	4,860
5	1,240	3	6,100
6	1,330	3	7,430
7	1,340	3	8,770
8	1,350	3	10,120
	31,870		

Stratum	Count	Value
1	22	10,830
2	10	10,920
3	8	10,120
	40	31,870

Proportional will attempt to have roughly equal recorded amounts in each stratum

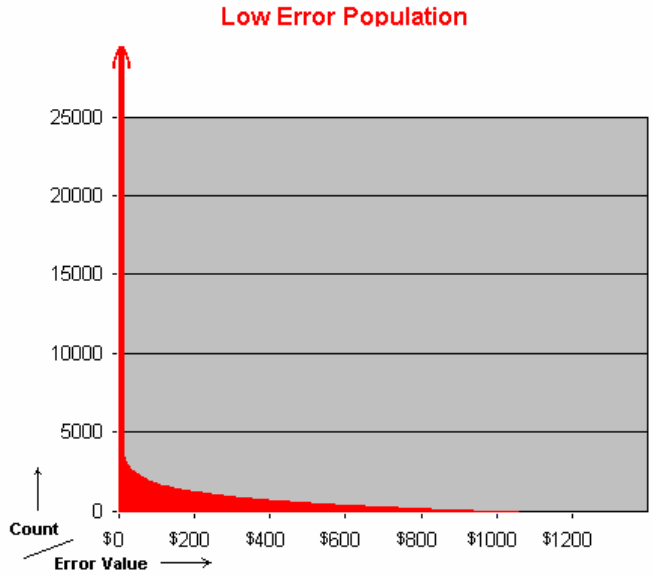
Geometric Ratio[2]

This method assumes a certain shape or distribution to taxable error values (y_i) that is not an exact fit, but often is close enough to provide efficiencies on the scale of the optimal method discussed in Appendix G. The method presumes a positive skew, and an exponential shape to the taxable error values. The more closely the taxable errors values follow the invoice values, the more likely this method will provide better efficiencies. A typical invoice population could look something like this:



An exponential distribution (population)

An error population could look like this:



[2] Introduced in an Abstract: *Contaduría y Administración*, 11 (September-December 2004), Universidad Nacional Autonoma de Mexico (UNAM), "Geometric Stratification of Accounting Data," by Patricia Gunning, Jane Horgan, and William Yancey.

These two diagrams are not the same, but have similarities. Also, individual accounting and taxable error populations can vary widely in shape and distribution. But typical accounting populations look like the exponential population in the first image. Taxable error populations that have no tax overpayments will often look like the second diagram. Of course, taxable error populations will often have mixtures of tax overpayments and tax deficiencies, as well as widely differing percentages of zero values. As the mixture of error types increases, or the percentage of non-errors (zero values) increases, this method will likely become less efficient.

Another potential problem with this method is that unlike CSRF and equal invoice amount methods, the Geometric Ratio method ignores the actual distribution of the invoice amounts between the smallest and largest value used in the formulas. This can often lead to odd strata break values that don't make a lot of sense. In some cases, the strata break values are not going to provide a workable stratification.

Another word of caution about the method needs to be mentioned. Proponents of the method tout its simplicity. It is in fact very simple when compared to the much more complex CSRF. However, it does not eliminate any of the other complex procedures, such as optimal allocation of the sample and statistical evaluation of the results. In fact, because of the problems mentioned above, the procedure is probably even more dependent on optimal sample allocation.

The mechanics of Geometric Ratio is straight forward. The auditor needs to establish a minimum and maximum invoice value. This can be often be done by sorting the population. The auditor also needs to decide on the number of strata sampled, or L. The minimum and maximum values are referred to as the floor and ceiling:

floor	k_0
ceiling	k_L

Using k_0 and k_L, a ratio, r, is calculated that is used to determine the strata break values:

$$r = \left(\frac{ceiling}{floor} \right)^{1/L}$$

The strata break values will be defined in the range of the population invoice values x_i:

$$k_0 < strata_break_values < k_L$$

The auditor needs to determine *L-1* strata break values k_h where *h* is from 1 to *L-1*: k_1, k_2 … k_{L-1}. The strata break values are determined according to the following formula:

$$k_h = k_0 * r^h$$