

## Cumulative Square Root of the Frequency (CSRF-U)<sup>i</sup>

Once the sampling frame is established, there are several ways to determine stratum breaks. In CAAP audits, the "Cumulative Square Root of the Frequency" is the preferred method to establish strata boundaries.

### Background

Stratification boundaries can be determined using several approaches. The most common method is CSRF. There are many references for this methodology.<sup>ii</sup> The CSRF methodology has been attributed to an article appearing in the Journal of American Statistical Association in 1959.<sup>iii</sup>

It would be optimal to establish strata boundaries based on the error amounts. Since the errors are unknown, populations are stratified using the invoice amounts.<sup>iv</sup> CSRF will distribute the variation of the invoice amounts in an "optimal" manner to the strata.

CSRF methodology breaks down the population into intervals. These intervals are sometimes referred to as "cells" which can be of equal or unequal width. In most references, the method is described using cells of equal width. When broken into unequal widths, an extra step is involved.<sup>v</sup> CAAP often uses cells of unequal width because the stratum breaks may become more uniform. Note that CSRF does not dictate the number of strata, only the breaks between the strata.

### Mechanics

The following are the procedures used to determine stratum breaks assuming cells of unequal width.

Step 1 Evaluate the population and determine if transactions below a certain dollar level should not be sampled based on the dollar significance of those transactions.

In this example, the taxpayer and the auditor agreed on the accounts of interest and the sample population. No transactions equal to or less than \$100.00 will be reviewed.

Step 2 Evaluate the population and determine the dollar amounts that will be reviewed on an actual basis.

All items greater than \$25,000.00 will be reviewed in detail (actual basis).

Step 3 Stratify the remaining transactions within the population of items to be sampled (\$100.01 - \$25,000.00) into dollar ranges.

Step 4 Determine the frequency for each dollar range. This is the number of transactions within each range.

- Step 5 Calculate the square root of the frequency (number of transactions) for the first range (\$100.01 – \$150.00). This value is 129.82. Then calculate the square root of the next range (\$150.01 – \$200.00). This value equals 101.01. Continue this process for each of the ranges.
- Step 6 Calculate the square root of the range width for the first range (\$150.00 – \$100.01). This value is 7.07. Then calculate the square root of the next range (\$200.00 – \$150.00). Again, this value equals 7.07. Continue this process for each of the ranges.
- Step 7 For each range; multiply the square root of the frequency (step 5) times the square root of the range (step 6).
- Step 8 Accumulate the values (as calculated in step 7) for each range. For the first range, the cumulative square root value is 917.87. The cumulative value for the second range is 1,632.15 (714.28 + 917.87 - accumulated value for the preceding range).
- For each range, add the square root value (step 7) to the preceding square root value (step 7) to arrive at the cumulative square root value. Continue this process for each of the ranges.
- Step 9 Divide the total cumulative square root value (18,422.66) by the number of strata (4 in this audit) to arrive at cumulative square root value for each stratum. In this audit the value is 4,605.67.
- Step 10 Multiply the result of step 9 by 1 to calculate the first stratum range, or 4,605.67. Locate the cumulative square root value closest to this amount. In this case, it is 4,623.62. Locate the ending range associated with this value (\$500.01 - \$550.00). The first stratum range (boundary) is \$100.01 – \$550.00.
- Step 11 Multiply the result of step 9 by 2 to calculate the second stratum range, or 9,211.33. Locate the cumulative square root value closest to this amount. In this case, it is 9,213.44. Locate the ending range associated with this value (\$2,100.01 – \$2,200.00). The second stratum range (boundary) is \$550.01 – \$2,200.00.
- Step 12 Multiply the result of step 9 by 3 to calculate the third stratum range, or 13,817.00. Locate the cumulative square root value closest to this amount. In this case, it is 13,871.18. Locate the ending range associated with this value (\$8,250.01 – \$8,500.00). The third stratum range (boundary) is \$2,200.01 – \$8,500.00.
- Step 13 Multiply the result of step 9 by 4 to calculate the fourth stratum range, or 18,422.66. Locate the cumulative square root value closest to this amount. In this case, it is

18,422.66. Locate the ending range associated with this value (\$20,000.01 – \$25,000.00). The fourth stratum range (boundary) is \$8,500.01– \$25,000.00.

Step 14 Calculate the coefficient of variation (standard deviation divided by the mean) for each of the strata (4 strata in this audit). Compute the average coefficient of variation for all the strata. An average coefficient of variation of less than 50% is the target range. In this example, the combined coefficient of variation is 40.20%.

Step 15 If the average coefficient of variation is greater than 50%; consideration should be given to adding another stratum. On the other hand, if the percentage is significantly less than 50%, then the total number of strata may be reduced.

### Example

Attached is an example of a workpaper applying the "Cumulative Square Root of the Frequency" method.

### Endnotes to Attachment:

<sup>i</sup> Taken from Washington State Department of Revenue's *Statistical Sampling Manual*, Appendix F.

<sup>ii</sup> The following are some of the references:

1. *Statistical Auditing*, Donald Roberts, AICPA, 1978, pages 97 to 98.
2. *Sampling Techniques*, William Cochran, John Wiley and Sons, 1977, pages 127-131.
3. *Applications of Statistical Sampling to Auditing*, Alvin Arens & James K Loebbecke, Prentice Hall, Englewood CA 1981, pages 256 – 258.
4. IRS training publication: *Advanced Statistical Sampling*, Training 3174-002 Rev (05-92) TPDS 87030A in Lesson 10, pages 6 – 10.
5. *Survey Sampling*, Leslie Kish, John Wiley & Sons, 1995, pages 104 –106.
6. *Sampling of Populations: Methods and Applications*, 3rd Edition, Paul Levy and Stanley Lemeshow, John Wiley & Sons, New York, 1999, pages 179 - 183.
7. *Elementary Survey Sampling*, 5th edition, Richard Scheaffer, William Mendenhall, R. Lyman Ott, Duxbury Press, New York, 1996, page 165.

<sup>iii</sup> According to *Elementary Survey Sampling* Scheaffer et al. footnote 13, page 189, CSRF was developed by T. Dalenius and J.L. Hodges Jr.

<sup>iv</sup> Generally, there is a relationship between error and invoice amounts. We know that the variation in the errors has something to do with the range of values in the non-zero error amounts. However, a larger contributor to the variation of errors may be the number of zero errors as percentage of the total (that is, the more zero errors, the more variable the population). The number of zero errors in the population will not have anything to do with variation in values of invoice amounts. CSRF will not consider the effect of the zeros or where nonzero errors include both debits and credits.

<sup>v</sup> In *Statistical Auditing*, Roberts describes the methodology of unequal lengths in a footnote on page 98. In the book by Cochran, *Sampling Techniques*, he describes this modification on page 130.