

Procedures for Data Stratification

A Simplified Procedure

XL Audit Commander

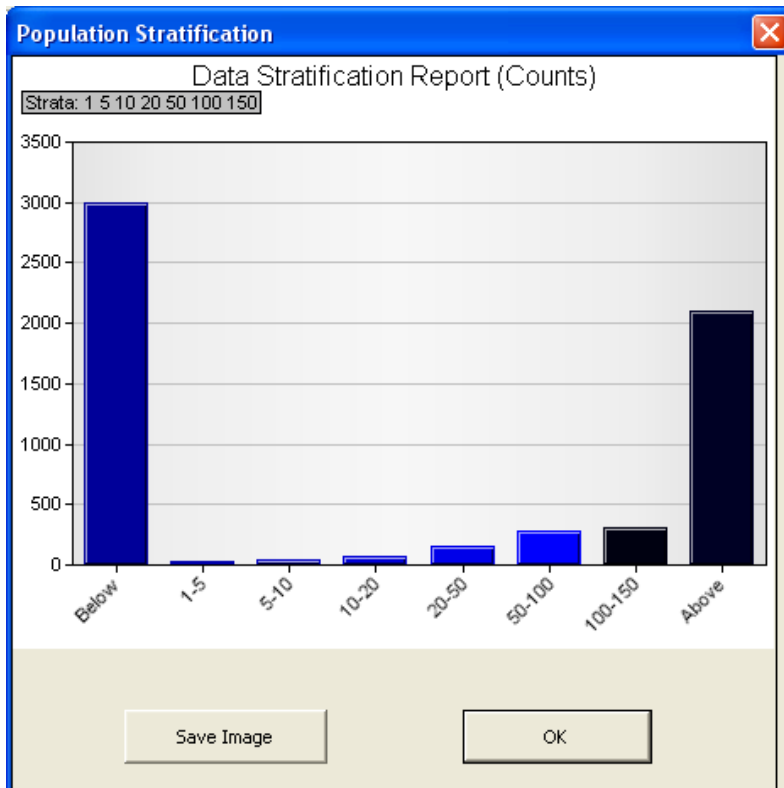
data analysis made easier ...

Data Stratification

Data stratification helps the auditor/researcher classify data, in order to perform tasks such as:

- Sample planning
- Audit Planning
- Identification of potential exceptions
- Determining if “outliers” may exist
- Obtaining a better understanding of the population
- Comparison of two components of one population

The procedures described here are an efficient and effective way to stratify the numeric data in the population being tested. . The procedure requires the use of the XL Audit Commander, a free tool available for download from <http://ezrstats.com/>. The tool is installed as an Excel add-in, and as such, requires Excel 2000 or later. The tool works only on Windows operating systems.



Data Stratification

Table of Contents

Table of Contents

Data Stratification	1
The stratification process	1
Typical audit areas	1
Classifying and Quantifying occurrences	2
Test data used.....	2
User Interface.....	2
Procedural Steps.....	5
Graphic Output.....	10
Related Areas of Interest.....	10
Summary and conclusion.....	12

Procedural steps

Data Stratification

The stratification process

Data stratification is a technique used to classify data into ranges. These ranges may consist of any values. Thus data stratification is a more general process than *histograms* which consist of bins of a uniform width, or the even more specialized percentiles and quartiles which classify data into specific percentages of the population.

Typical audit areas

Most often data stratification is used for audit planning and sample design. However, there are many areas where stratification may be useful, including:

- Inventory counts
- Determination of possible unusual items or exceptions
- Obtaining an understanding of the most common values
- Testing Accounts receivable balances
- Reviewing Journal entries

Procedural steps

Classifying and Quantifying occurrences

Data stratification can be done using formula in Excel. For example counts may be obtained using either the COUNT function or the COUNTIF function to consider conditions. Similarly, the SUM and SUMIF formula can be used to obtain amounts. The standard Analysis Tool Addin also has some capabilities to provide histograms.

However, a most common approach to data stratification is the iterative approach, i.e. specify some strata and review the results obtained. Then, it is often desirable to redefine the strata and obtain new results. This process may be repeated numerous times until the auditor/researcher is satisfied with the results.

Test data used

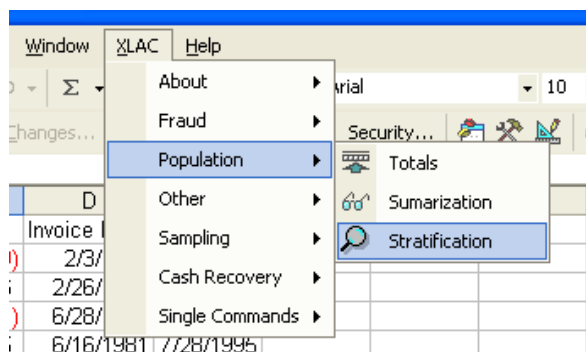
The data used in this article is available in the Excel workbook located at <http://ezrstats.com/online/inno/QS.xls>. The sheet named “t_CMADData” contains hypothetical invoice amounts to be tested.

User Interface

The tool can be used in a combination of four modes:

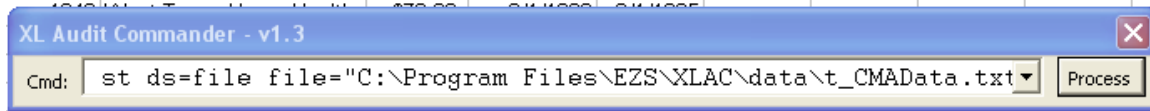
1. Menu – a graphical system to gather the required information
2. Command Bar – commands are typed as text
3. System registry – commands are stored in the system registry and then loaded for selection, modification and processing
4. \$Cmd worksheet – the main processing parameters are stored on a worksheet

The menu system allows the necessary data to be gathered in a graphical manner. Once the data has been gathered, the “Process” button is clicked and the commands are then written to the toolbar for processing. An example of the main menu is shown below (Note the added menu option “XLAC” which appears at the top of the menu bar).



Procedural steps

To perform the testing, the needed information can be either entered directly using the menu or else typed into the toolbar, which is show below.



Note that if the menu system is used, then the results will be to translate the menu information into an audit command which will then be copied into the toolbar where it can then be processed.

Data stratification can be performed on four types of data sources:

1. Data contained on a sheet in an Excel workbook
2. Data contained within a highlighted range of an excel sheet
3. Data contained in text file which is in tab separated value format
4. Data contained in a database

The first page of the test data being used for illustration in this article is shown below. This data represents hypothetical invoice details. The column “Amount” will be used as the data to be stratified.

1	Check Number	Payee	Amount	Invoice Date	Due Date
2	2985	Shreiner, Marla Rae	(\$234.49)	2/3/1959	8/22/2000
3	1148	Miesz, Andrew	\$460.56	2/26/1978	7/31/1995
4	1543	Sheikholeslam	(\$163.51)	6/28/1978	8/1/1995
5	821	Hutchens, Jerome	\$115.75	6/16/1981	7/28/1995
6	1802	Watley, Vernon Thompson	(\$36.21)	11/10/1981	8/1/1995
7	935	Kones, Richard T	(\$149.47)	2/26/1982	7/28/1995
8	1317	Peters, Andrew	(\$120.29)	4/12/1982	8/1/1995
9	1739	Tyler Communica	(\$205.51)	4/30/1982	8/1/1995
10	788	Houston Child Gu	\$100.19	7/19/1982	7/28/1995
11	1819	Weinheimer, E A	(\$16.00)	8/19/1982	8/1/1995
12	1831	Wess, Mitchell O	(\$468.30)	9/15/1982	8/1/1995
13	792	Howel, Larry	\$160.74	2/28/1983	7/28/1995
14	791	Howels, Super Duper	(\$469.51)	2/28/1983	7/28/1995
15	1121	Medical & Surgical	\$211.30	3/1/1983	7/31/1995
16	1177	Moore, Dudley	(\$494.27)	4/12/1983	7/31/1995
17	1214	Naifeh, Jerome G	(\$184.80)	7/28/1983	8/1/1995
18	1838	West Texas Ho	(\$175.22)	9/1/1983	8/1/1995
19	1835	West Texas Home Hea	\$338.71	9/1/1983	8/1/1995
20	1837	West Texas Home Heal	(\$127.11)	9/1/1983	8/1/1995
21	1836	West Texas Home Health	(\$279.13)	9/1/1983	8/1/1995
22	1839	West Texas Home Health	\$422.56	9/1/1983	8/1/1995
23	1840	West Texas Home Health	\$174.05	9/1/1983	8/1/1995

Procedural steps

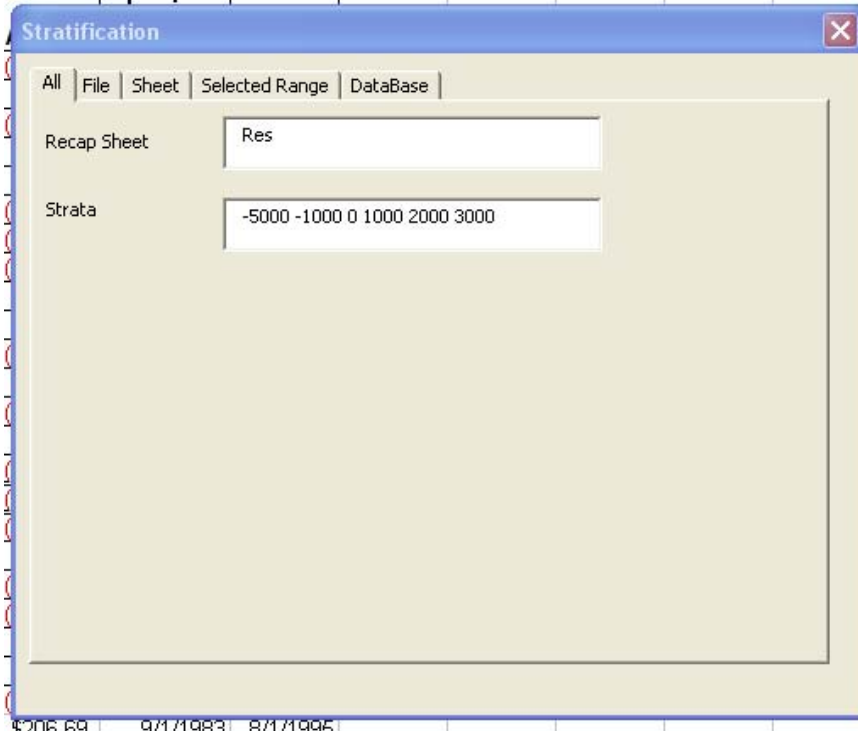
The first test to be performed is to stratify the amounts in the column named “Amount”. The values on the worksheet named “t_CMADData” will be tested starting at the upper left hand corner, i.e. cell “A1” and continuing down to the first blank cell encountered.

For illustration purposes, both the menu options as well as the toolbar commands will be described. First select the first tab labeled “All” and specify the name of the worksheet to contain the results of the analysis. In this case, a new worksheet named “Res” (results) will be specified. (Res is the default name). The initial strata to be used is -5000, -1000, 0, 1000, 2000, 3000. Note that commas are not included in the specification of the strata.

Procedural steps

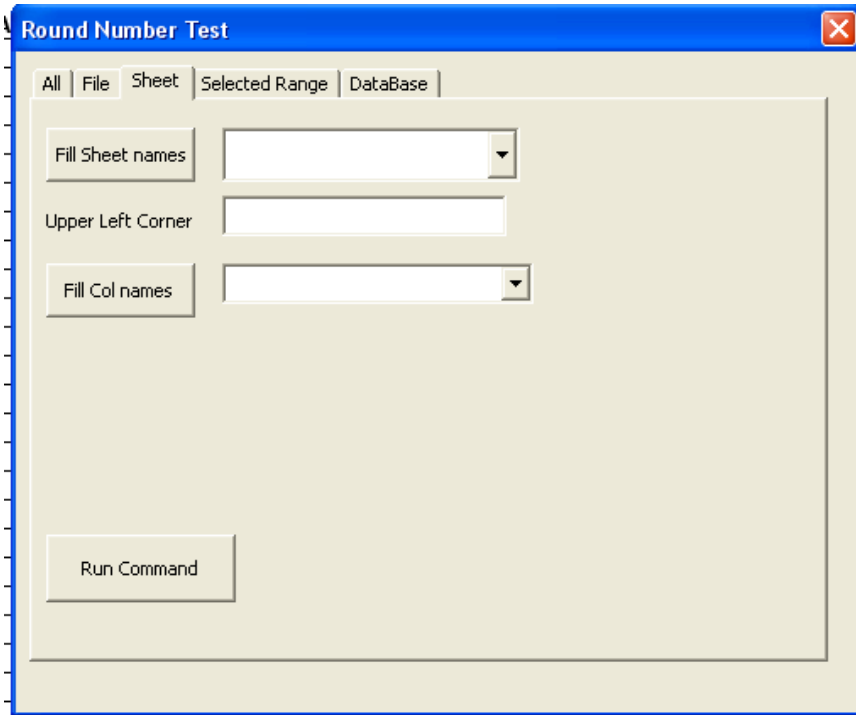
Procedural Steps

Step by step example

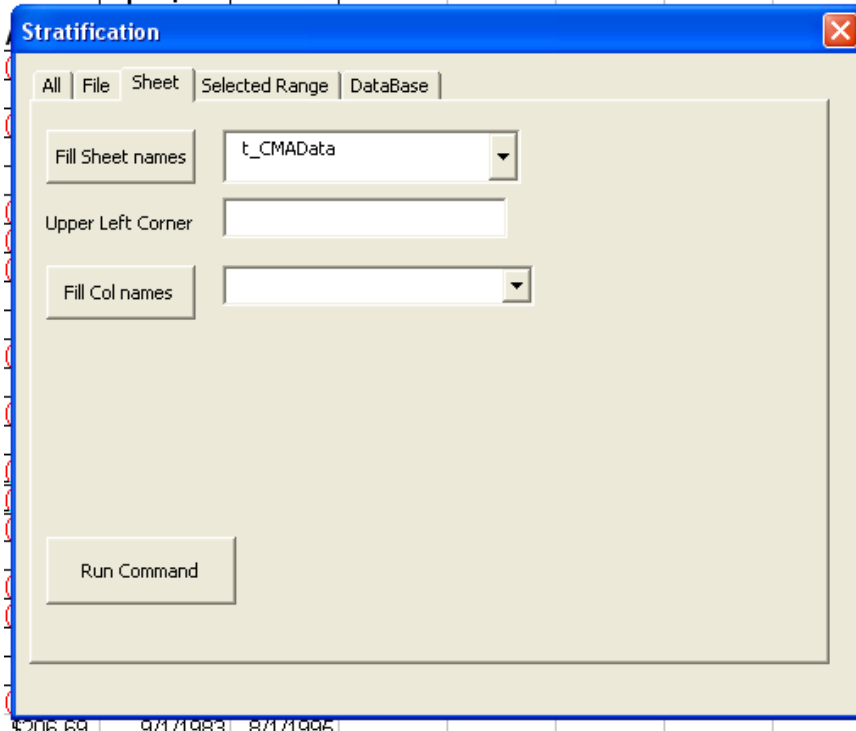


Now click on the tab labeled “Sheet” in order to specify the sheet name, upper left hand corner and column to be analyzed:

Procedural steps

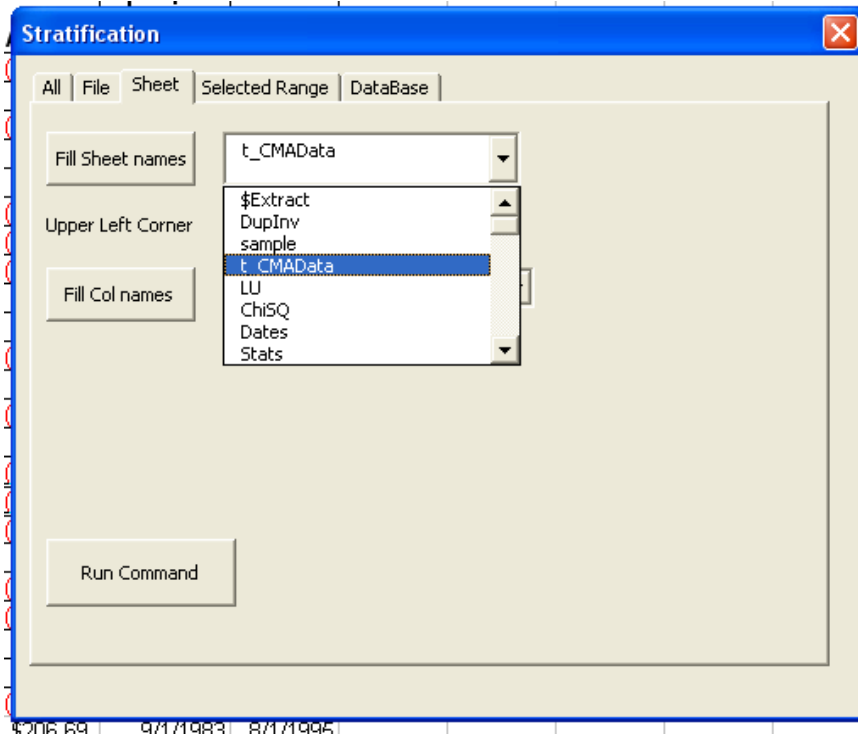


First, have the system fill in all the sheet names into the first drop down list box. This is done by clicking on the button labeled “Fill sheet names”:

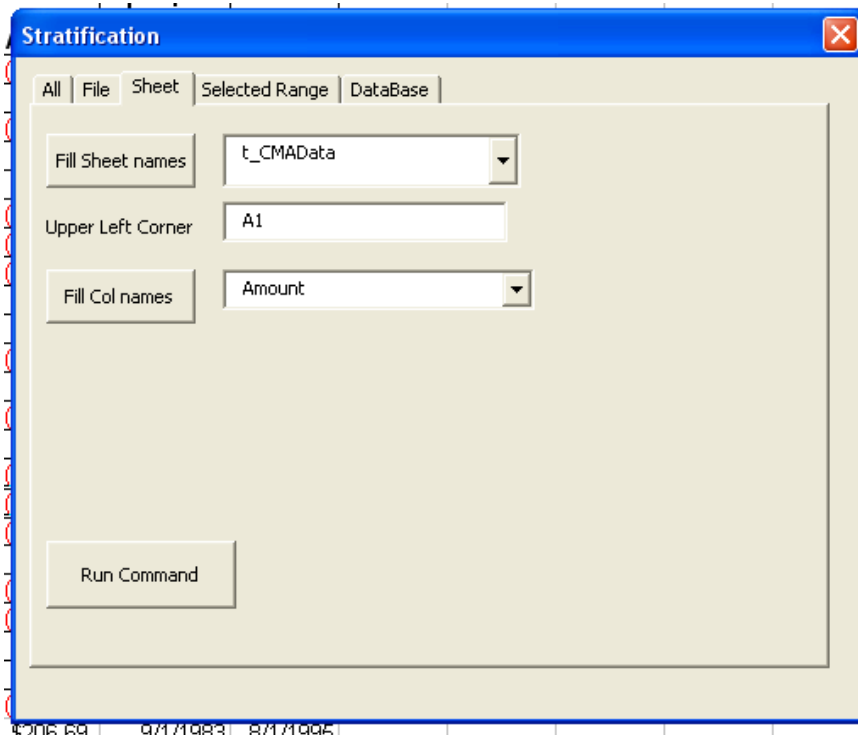


Now, click on the drop down list to select the sheet named “t_CMADData”:

Procedural steps

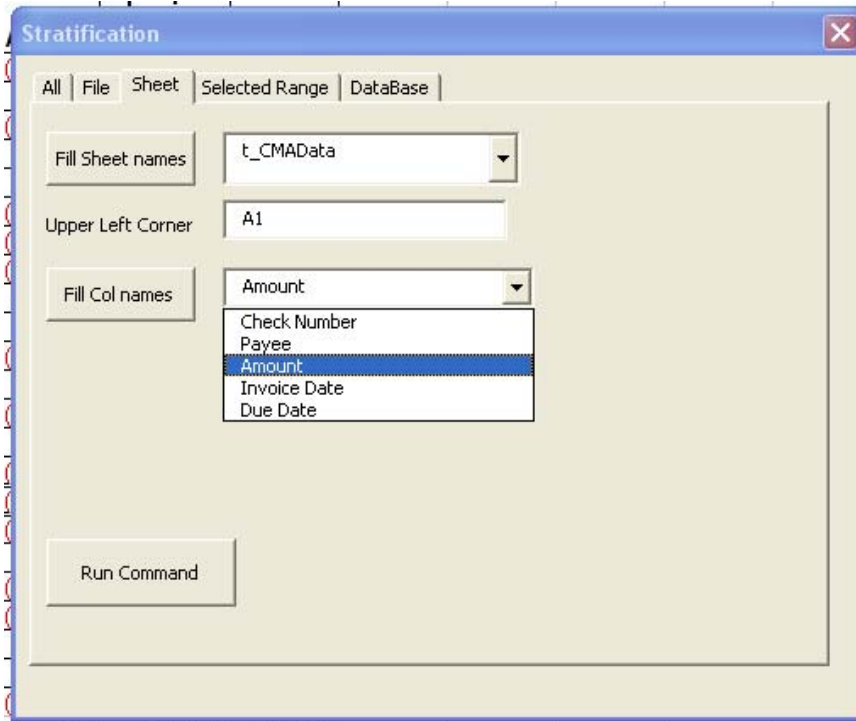


Select “t_CMADData” and then enter the cell of the upper left hand corner, which is “A1” in this instance:

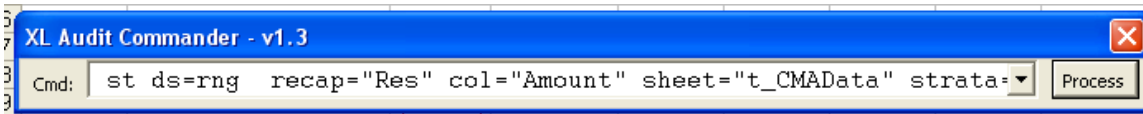


Procedural steps

Now the system can fill in all the column names, in order that the numeric column to be tested can be specified:



The next step is to click the “Run Command” button, in order to construct the command to be placed into the toolbar. The results are shown below:



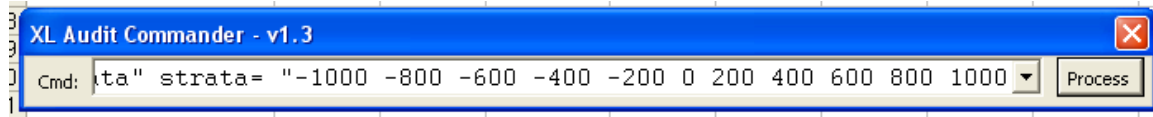
Then click on the “Process” button in order to perform the analysis which is stored on the sheet named “Res” (if it doesn’t exist, it will be created, if it does exist it will be overwritten).

Summary for Strata		-5000	-1000	0	1000	2000	3000
Start	End	Count	Amount	Pct	Cumulative		
Below	Below	0	0	0	0		
	-5000	-1000	0	0	0		
	-1000	0	2,985	-747,720.32	159.5743	159.5743	
	0	1000	2,988	743,034.60	-158.574	1	
	1000	2000	0	0	0	1	
	2000	3000	0	0	0	1	
Above	Above	0	0	0	0	1	
Totals	totals	5,973	-4,685.72				

Procedural steps

This report is not that helpful, as all the values are clustered between -1000 and 1000. Therefore, different strata must be specified. We will start with uniform amounts with a width of 200, i.e. -1000 -800 -600 -400 -200 0 200 400 600 800 1000.

The changed strata can be typed directly into the toolbar (no need to use the menu again).



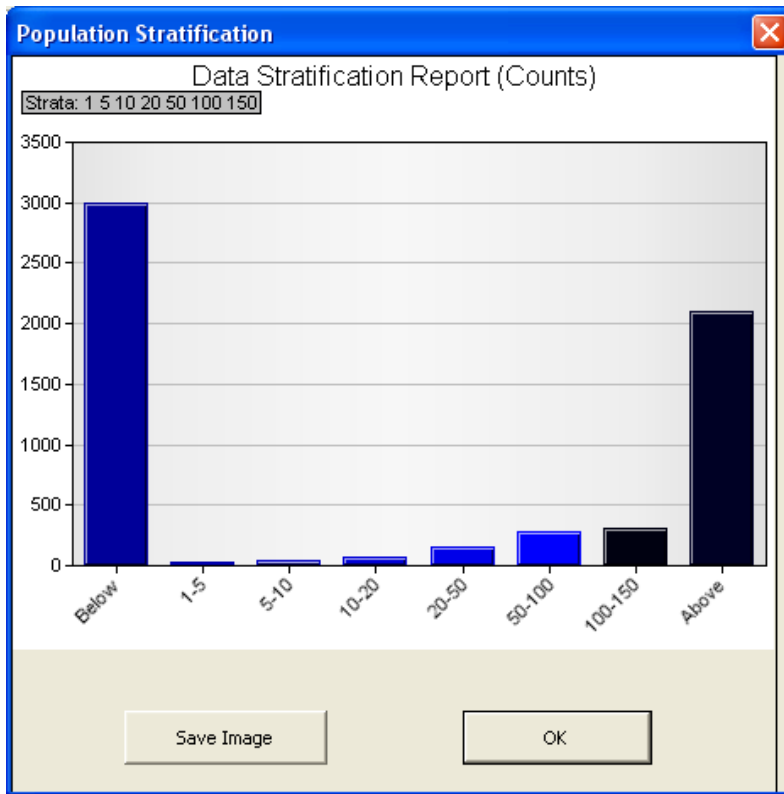
Clicking the “Process” button yields the following results:

Summary
for Strata -1000 -800 -600 -400 -200 0 200 400 600 800 1000

Start	End	Count	Amount	Pct	Cumulative
Below	Below	0	0	0	0
-1000	-800	0	0	0	0
-800	-600	0	0	0	0
-600	-400	615	-277,044.12	59.1252	59.1252
-400	-200	1,158	-349,889.96	74.6715	133.7967
-200	0	1,212	-120,786.24	25.7775	159.5743
0	200	1,171	116,922.83	-24.953	134.6213
200	400	1,237	366,820.89	78.2849	56.3364
400	600	580	259,290.88	55.3364	1
600	800	0	0	0	1
800	1000	0	0	0	1
Above	Above	0	0	0	1
Totals	totals	5,973	-4,685.72		

Procedural steps

Graphic Output



Related Areas of Interest

Other related documents/guides of possible interest include:

Topic	Description
Auditing Data in Access	An easier way to perform 18 audit tests on data in Microsoft Access® http://ezrstats.com/online/AuditGuide/Auditing_Data_in_MS_Access_Databases.pdf
Auditing Data in Worksheets	18 audit tests for data stored in Excel worksheets http://ezrstats.com/online/AuditGuide/Auditing_Data_in_Workbooks.pdf
Auditing Data in Files	18 audit tests to perform on data files in tab separated value format http://ezrstats.com/online/AuditGuide/Auditing_Data_in_Files.pdf

Procedural steps

Round Numbers	Why to check for "round" numbers and how
http://ezrstats.com/online/AuditGuide/Testing_For_Round_Numbers.pdf	
Holidays	Identification of holiday dates, e.g. in Journal entries, invoices, etc.
http://ezrstats.com/online/AuditGuide/Testing_For_Holidays.pdf	
Data Stratification	Stratification as a planning and audit tool
http://ezrstats.com/online/AuditGuide/Procedures_For_Data_Stratification.pdf	
Cross tabulations	Use of cross tabulations in audits
http://ezrstats.com/online/AuditGuide/Cross_Tabulations_As_An_Audit_Technique.pdf	
Benford's law	Test conformity with Benford's Law
http://ezrstats.com/doc/Auditors_Guide_to_Tests_using_Benford's_Law.pdf	
Basic Data Extraction	Extracting data based upon criteria, and performing calculations
http://ezrstats.com/online/AuditGuide/Basic_Data_Extraction_Techniques.pdf	
Data Classification	Basic techniques for classifying data Software Installation
http://ezrstats.com/online/AuditGuide/Basic_Data_Classification_Procedures.pdf	
Setup.exe	Setup file - double click to install (6.0 MB)
http://ezrstats.com/online/inno/XLACSetup.exe	
Install Instructions	Installation Guide (PDF document) (.7 MB)
http://ezrstats.com/online/inno/XL_Audit_Commander_Installation_Guide.pdf	
Operation Guide	Operation Guide (PDF document) (2.5 MB)
http://ezrstats.com/online/inno/XL_Audit_Commander.pdf	
Quick Start	Quick Start Module (Excel Workbook - open after install) (3.1 MB)
http://ezrstats.com/online/inno/QS.xls	
Help	Shows list of help links in the current workbook
http://ezrstats.com/helpxlac/he.php	
Single Commands	Commands of just two letters for a selected range on a single worksheet
http://ezrstats.com/helpxlac/single.php	
Population	Population statistics (univariate, stratify, population, duplicates)
http://ezrstats.com/helpxlac/ndxpop.php	
Sampling	Sampling procedures (cma, interval, sample size calculation, precision calculation)
http://ezrstats.com/helpxlac/ndxsamp.php	
Fraud	Fraud investigation tools (test Benford's Law, duplicates)
http://ezrstats.com/helpxlac/ndxfraud.php	
Cash Recovery	Cash Recovery procedures ("Near miss" invoices, split invoices)
http://ezrstats.com/helpxlac/ndxcr.php	

Procedural steps

Other	Other Commands (ageing, gaps, credit card validation, analytic review procedures, dates on federal holidays, etc.)
http://ezrstats.com/helpxlac/ndxoth.php	

Summary and conclusion

Data stratification should be a common audit procedure, especially when the data to be tested has already been converted to electronic format. The procedure should not take long, and may yield some interesting results for follow-up!

There are a host of other audit procedures which may also be of interest. There is an index to the procedures at <http://ezrstats.com/helpxlac/index.php>. Software and documentation can be downloaded from <http://ezrstats.com/online/inno/XLACSetup.exe>, and from http://ezrstats.com/online/inno/XL_Audit_Commander.pdf.