Department of Health & Human Services
OIG - Office of Audit Services

RAT-STATS

USER GUIDE

September 2001

TABLE OF CONTENTS

SAMPLE SIZE DETERMINATION

Variable Sample Size Determination - Stratified

Attribute Sample Size Determination

INTRODUCTION

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
RAT-STATS
VERSION: September 2001

RAT-STATS is a package of statistical software tools to assist the user in performing random samples and evaluating the results. The package is designed to run on personal computers using Microsoft Windows (Windows 95 and later versions).

We have attempted to make the software as "user friendly" as possible, keeping in mind the user is working with the technical terms of statistics. We have also attempted to make the software flexible in terms of entering data and allowing output to a variety of devices.

Installation

The executable version of this program is called **RATSTATSV1**. This program can be installed by double clicking on **Setup.exe**. After installing RATSTATSV1, the program can be executed by double clicking on this file. An easier procedure would be to install a shortcut to RATSTATSV1 on the windows desktop.

This guide contains explanations and examples for using each of the programs. To understand how to use any particular option, look up the program in the table of contents and proceed to the appropriate pages for an overview, explanations and examples.

Frequently Asked Questions

ARE THE RANDOM NUMBERS REALLY RANDOM?

The random number generator used throughout this package is based on an article entitled, "Building a Random Number Generator" that appeared in the March 1987 issue of Byte magazine (pages 127 and 128). The random number software in this package was tested with thirteen certification programs from the National Bureau of Standards to test for various aspects of randomness. The software passed all thirteen of the tests.

WHAT IS THE COMPUTER LANGUAGE OF RAT-STATS?

The software was written using Microsoft Visual Basic (Version 6).

WHO IS RESPONSIBLE FOR RAT-STATS?

Several individuals have been involved in the development of this package. Key members for this edition are:

Janet Fowler, PhD - Statistician, Office of Audit Services
Al Kvanli, PhD - Associate Professor, Department of Business Computer Information Systems, College of Business Administration, University of North Texas
Doug Rennie - Manager, RATS - San Francisco

WHOM DO I CONTACT ABOUT RAT-STATS?

If you have questions about the software or suggestions for improvements, you may contact Doug Rennie on the Internet at drennie@oig.hhs.gov.

WHY IS IT CALLED RAT-STATS?

This is the most common question asked about the package. The software was initially developed by the Regional Advanced Techniques Staff (RATS) in San Francisco. After an informal naming contest in the RATS office, the name **RAT-STATS** won by a process of elimination.

WILL THERE BE FUTURE UPDATES?

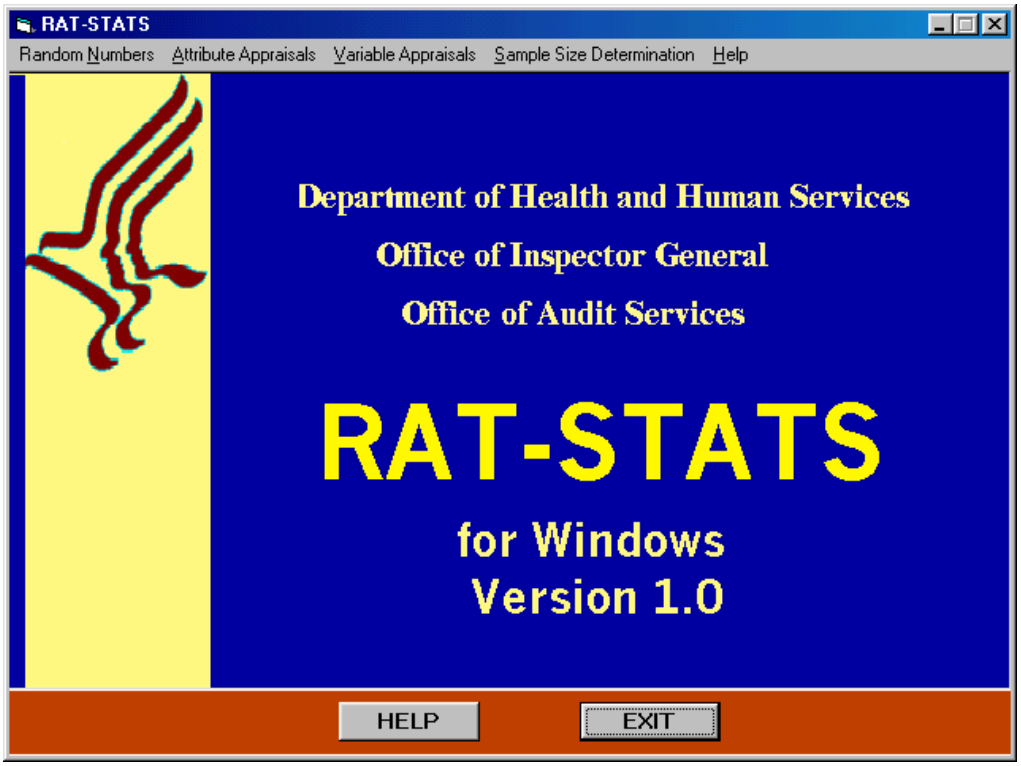
We are presently expanding this version of RAT-STATS to include all the modules contained in the DOS version.

Windows Version of RAT-STATS

RANDOM NUMBERS
ATTRIBUTE APPRAISALS
VARIABLE APPRAISALS
SAMPLE SIZE DETERMINATION

Opening Screen

The opening screen for the windows version of RAT-STATS is shown below. Click on **HELP** to view the RAT-STATS Help file and **EXIT** to exit the program.



Random Numbers

The RAT-STATS package contains a random number generator that should be used to randomly select items for review. Section two of this guide explains how to use the random number program.

Attribute Appraisals

Attribute appraisals allows a user to estimate the rate of occurrence of a given condition. The user may, for example, want to know the rate of occurrence of checks issued for less than $100. After drawing a random sample and evaluating the items selected, an attribute appraisal would be used not only to estimate the rate of occurrence, but also to determine (with a measured degree of confidence) the boundaries of the estimate.

By selecting **ATTRIBUTE APPRAISALS**, the user will see a window appear on the screen with the two attribute appraisal programs. Section three of this guide explains each of the attribute appraisal programs.

Variable Appraisals

The purpose of using variable appraisals is to measure a quantitative characteristic or set of characteristics. The user may, for example, want to know the value of all checks approved by a certain supervisor. After drawing a random sample and identifying the checks approved by the supervisor, a variable appraisal would be used not only to estimate the total value, but also to determine (with a measured degree of confidence) the boundaries of the estimate.

By selecting **VARIABLE APPRAISALS**, the user will see a window appear on the screen with the two variable appraisal programs. Section four of this guide explains each of the variable appraisal programs.

Sample Size Determination

Section five deals with sample size determination. The Variable Sample Size Determination program allows the user to estimate sample sizes for specified precision percentages and specified confidence levels. In the Variable Unrestricted module, the user will have the option of having the program read a probe sample file to obtain an estimate of the universe mean and standard deviation or input these two estimates directly without reading a probe sample file. The Variable Stratified module will determine optimum sample sizes for situations where the total sample size is both predetermined or unknown.

The Attribute Sample Size Determination program determines the sample size for an attribute simple random sample. The sample size is determined to provide for a specified degree of precision (using the desired width of the confidence interval) at four levels of confidence (80%, 90%, 95%, and 99%). The resulting sample sizes are the smallest sample sizes capable of meeting the specified precision requirement at the stated confidence level.

Random Numbers

OVERVIEW

SINGLE STAGE NUMBERS

Selecting items for a sample usually requires assigning a unique identifier to each item in the universe. Statistical textbooks typically assign a numbering sequence from one to the size of the sampling frame for their examples and problems. However, in the real world of sampling, the person drawing the sample may not find such a numbering sequence and the size of the frame may not make it feasible to manually create such a sequencing technique.

Using books of random digits may cause further delays in selecting the sample items. If, for example, the universe is numbered from 1 to 3000, approximately 70% of the 4-digit numbers drawn would have to be rejected since they fall outside the universe boundaries. In addition, the person drawing the sample would probably want to sample without replacement. By selecting this approach, the person must eliminate any duplicate selections of random numbers. This is usually accomplished by sorting the random numbers in ascending order and identifying duplicates in the process. The sorting of the random numbers process will usually speed up the sample selection process.

This package attempts to minimize the efforts of the user in actually identifying the items to be sampled. Once the user has identified the boundaries of the sampling frame, the random number module will ensure that only an unduplicated list of random numbers within the stated range is selected. The random number module also allows the user to have all or a portion of the random numbers sorted in ascending order to allow for a more efficient retrieval of the sampled items. The user also has the option of outputting the random numbers to any combination of the following output formats: Printer, text file, Access table, Excel spreadsheet, or flat file.

Single Stage Numbers

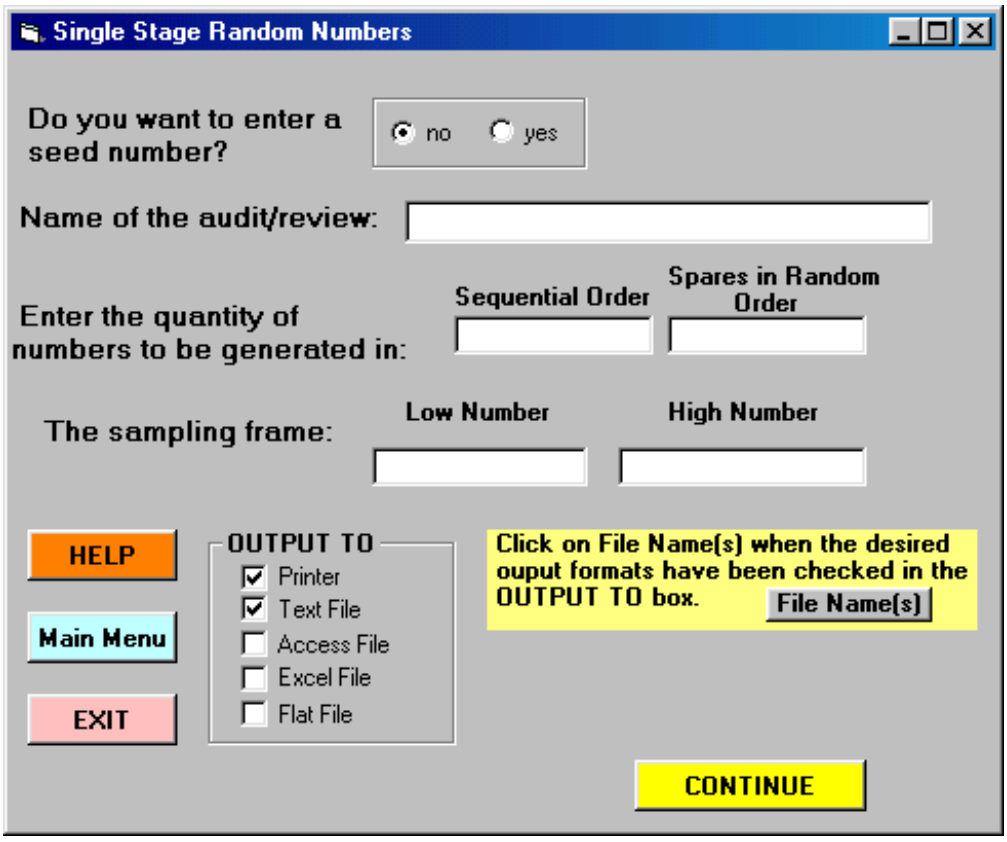
This module is used when the sampling frame of items has a numbering scheme. The frame could be a computer listing with each item numbered from one to the size of the sampling frame. The numbering scheme could also be based on check or voucher numbers assigned to each document.

SINGLE STAGE NUMBERS

Purpose

This program will generate an unduplicated quantity of random numbers (maximum of 2,000) for the user. The quantity of random numbers requested must be less than the size of the sampling frame.

Input Screen



NOTE: Example is for illustrative purposes only. The sample size may not conform with the organization's minimum sample size standards.

Do you want to enter a seed number?

The program allows a seed number to be entered by the user to start the random number generation. If no number is entered, then the program will use an algorithm to generate its own seed number. The seed number algorithm is based on the clock in the computer. If the computer clock is not functioning, the user should obtain and document a seed number from another source (e.g., a book of random numbers). The seed number will be printed as part of the output.

Name of the audit/review:

The user may enter a brief description to document the purpose of the run. The user's response will be placed at the top of each printed page. The description should be less than 40 characters in length and may include commas and spaces.

Enter the quantity of numbers desired to be generated in sequential order:

The quantity of random numbers to be generated in sequential order should be entered in this box. After the quantity indicated has been generated by the program, the random numbers will be sorted and the output will be in ascending order to assist the user in retrieving the sample items. The order of selection will be printed with the random numbers. If the quantity desired is zero, then a "0" (zero) must be entered.

Spares in random order:

The quantity of numbers to be generated in random order should be entered in this box. The random numbers will be displayed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

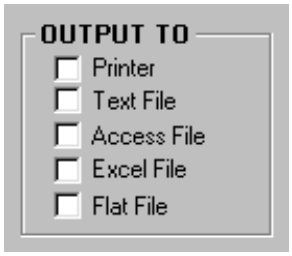
Entering the sampling frame

The low and high numbers in the sampling frame are the boundaries of the frame from which the user will be sampling. If the frame is a computer listing numbered 1 through 1,000 then the low entry will be 1 and the high entry will be 1,000. If the frame is a check register with checks numbered between 1,346 and 2,785, then the low will be 1,346 and the high will be 2,785.

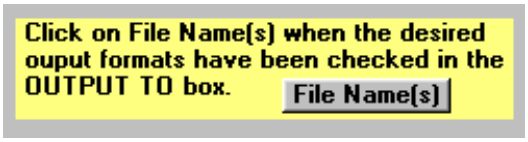
Output Options

The program allows for five output options. The user may select the output to be sent to printer, text file, access table, excel spreadsheet, or flat file. The user selects the appropriate output. The program always concludes with a summary on the screen.

Program Output

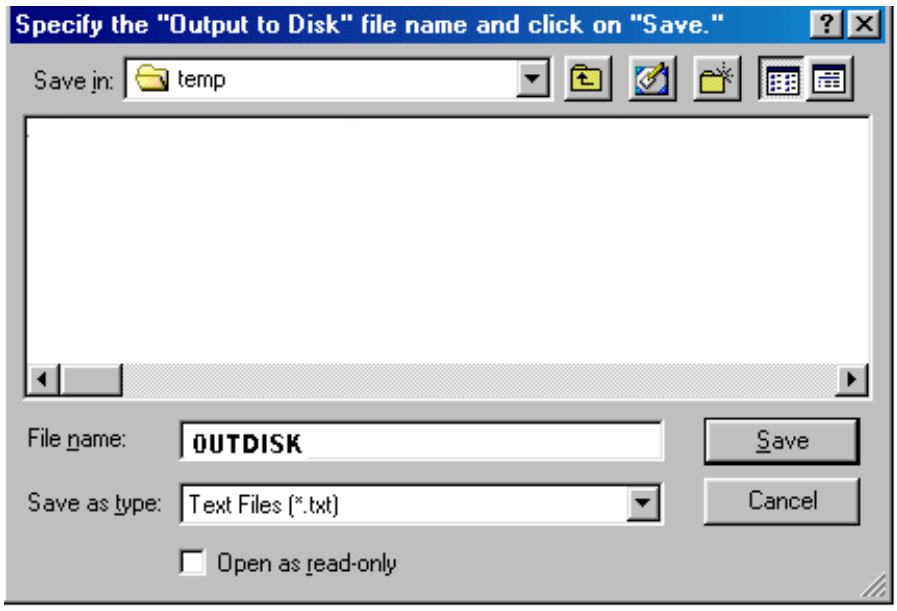


The program output can be to any combination of the above five formats. To select one or more output formats, select the corresponding device/file in the above OUTPUT TO list. If Text File, Access File, Excel File, or Flat File is selected, the following message will appear.



Output to a Text File

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. A text file will be saved with a ".txt" extension (e.g., C:\TEMP\OUTDISK.txt). Next, the user will see the standard Windows "Save" screen. Fill in the name of the file in the **File name** box. The file name for this illustration is C:\TEMP\OUTDISK.txt.



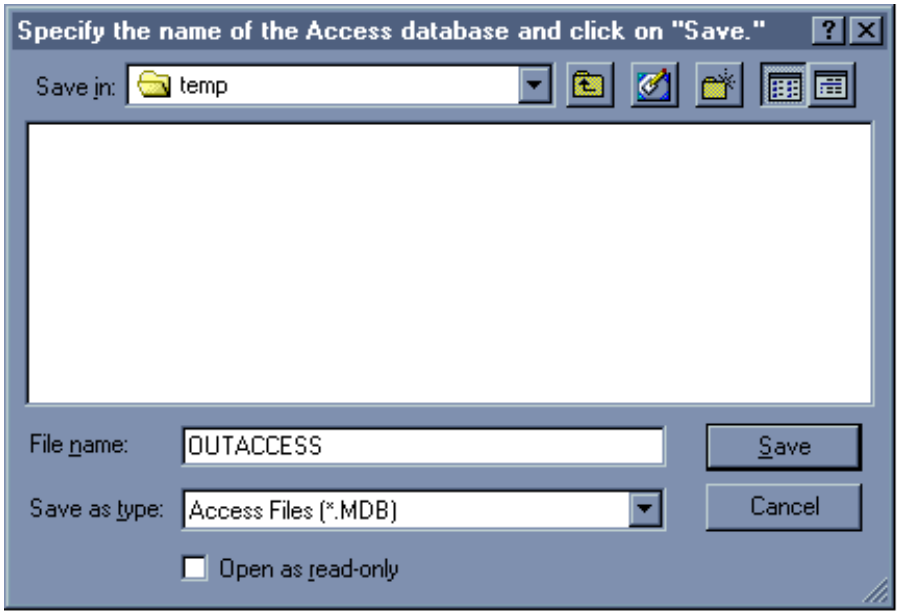
By clicking on the **Save** button, the user will return to the original input screen for this module. The output shown next is file C:\ TEMP\OUTDISK.txt and is the result of generating 10 random numbers between 1 and 1000 along with four spares. For each random number generated, two pieces of information are provided. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The second piece of information is the actual random number selected.

```
                Department of Health and Human Services
                    OIG - Office of Audit Services
Date: 12/16/1999      Random Number Generator          Time: 12:06
                         AUDIT: Example

SEED NUMBER: 42702.26                          FRAME SIZE:    1,000

         FILE OF RANDOM NUMBERS: C:\temp\OUTDISK.txt

         TOTAL RANDOM NUMBERS GENERATED: 14

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:
   POSITIONS 1 THROUGH  6 - ORDER OF SELECTION
   POSITIONS 7 THROUGH 17 - RANDOM NUMBER
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.
SUMMATION OF RANDOM NUMBERS = 7,461
```

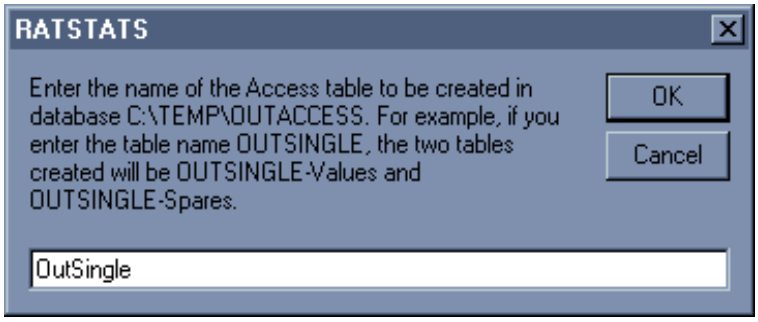
Selection Order	Value	
10	50	µ **The 10 random values start here.**
9	321	
3	359	
8	389	
2	484	
4	490	
1	491	
7	569	
5	733	
6	815	
11	599	µ **The 4 spare values start here.**
12	648	
13	921	
14	592	

Output to an Access Database

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. The user will then see the following standard Windows "Save" screen if the **OUTPUT TO Access file** is selected. Fill in the name of the Access database in the **File name** box. The name of the database for this illustration is C:\TEMP\OUTACCESS.MDB.



By clicking on the **Save** button, the user will return to the original input screen for this module. After clicking on **Continue**, the user will see the following form.



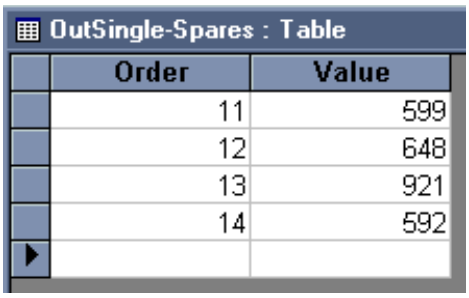
The program is asking for the name of the Access table to create in the specified database (C:\TEMP\OUTACCESS.MDB for this illustration). The table name "OutSingle" will be used. The program will then create two tables (OutSingle-Values and OutSingle-Spares) within database C:\TEMP\OUTACCESS.MDB.

The table OutSingle-Values shown next is the result of generating 10 random numbers between 1 and 1000 along with four spares. For each random number generated, two pieces of information are created. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The second piece of information is the actual random number selected. The seed number used by the random number generator is in the far-right column.

OutSingle-Values : Table

Order	Value	Seed-Number
10	50	42702.26
9	321	
3	359	
8	389	
2	484	
4	490	
1	491	
7	569	
5	733	
6	815	

The next table shows OutSingle-Spares containing four spare values.

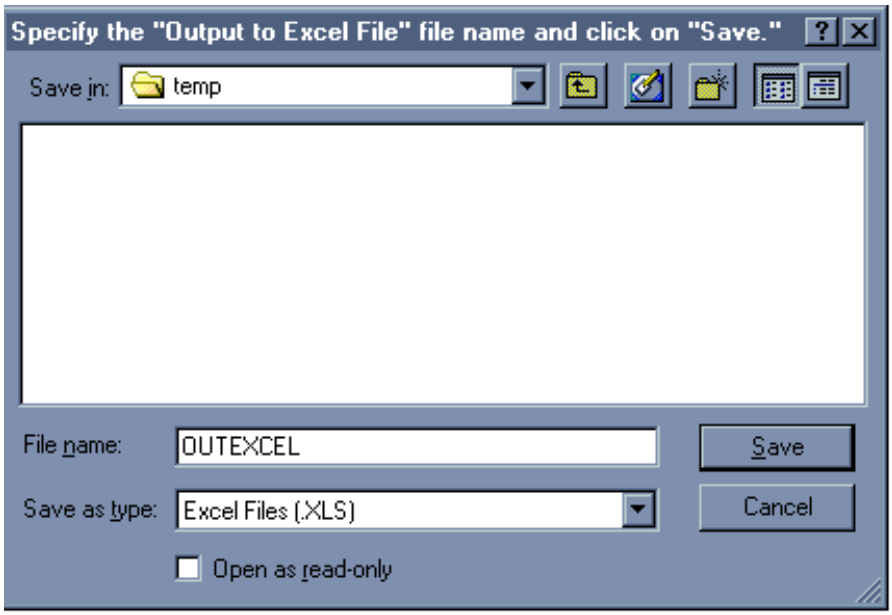


OutSingle-Spares : Table

Order	Value
11	599
12	648
13	921
14	592

Output to an Excel Spreadsheet

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. The user will then see the following standard Windows "Save" screen if the **OUTPUT TO Excel file** is selected. Fill in the name of the Excel workbook in the **File name** box. The name of the workbook for this illustration is C:\TEMP\OUTEXCEL.XLS.



If this workbook already exists, the user will be queried as to whether to replace the existing file. If the user selects "No," the program returns to the above form.

The following output will be contained in the Excel workbook C:\TEMP\OUTEXCEL.XLS. By clicking on the tab labeled "SPARES" the user will obtain the output immediately following. For each random number generated, two pieces of information are created. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The second piece of information is the actual random number selected. The output also contains the program execution date and time, the name of the audit/review, the seed number, and the size of the frame. **NOTE**: The user must first exit RAT-STATS in order to view this file.

	A	B	C	D
1	**Department of Health and Human Services**			
2	**OIG - Office of Audit Services**			
3	**Random Number Generator**			
4	**Date:**	12/16/1999	**Time:**	8:36
5	**Audit:**	example		
6	**Order**	**Value**	**Seed Number**	**Frame Size**
7	10	50	42702.26	1,000
8	9	321		
9	3	359		
10	8	389		
11	2	484		
12	4	490		
13	1	491		
14	7	569		
15	5	733		
16	6	815		

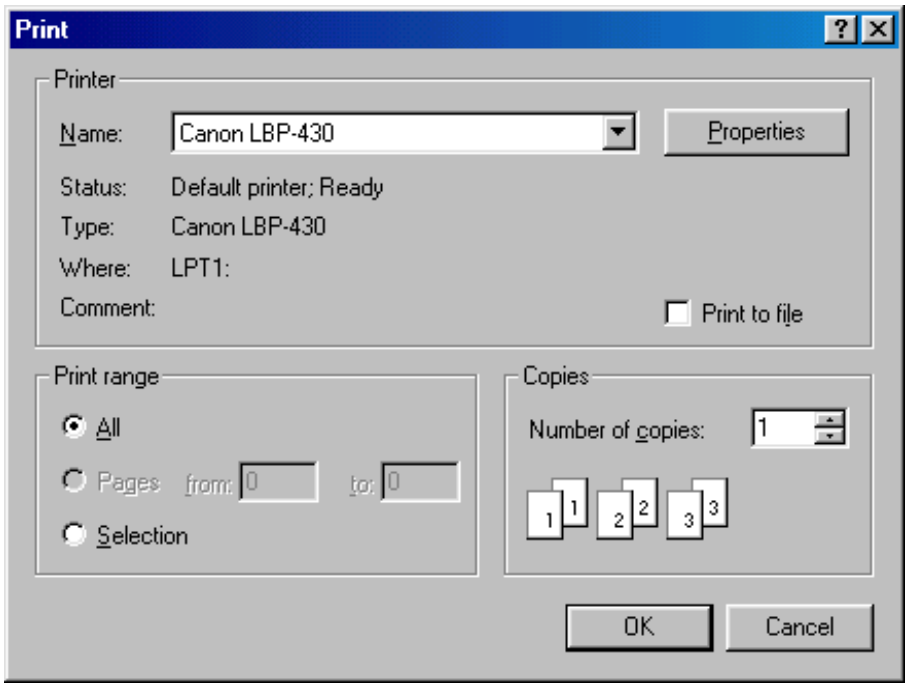
These are the ten random values in the Excel spreadsheet named "VALUES" located in C:\TEMP\OUTEXCEL.XLS

	A	B
1	**Order**	**Value**
2	11	599
3	12	648
4	13	921
5	14	592
6		

These are the four spares in the Excel spreadsheet named "SPARES."

Output to a Printer

If the user selects the printer for output, the standard Windows "Print" dialog box (shown next) will appear. Select the appropriate printer and click on **OK**.



The output immediately following is the printer output when generating 10 random numbers between 1 and 1000 along with four spares. For each random number generated, two pieces of information are displayed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The second piece of information is the actual random number selected. The subsequent random numbers are in sequential order going from left to right across the page. Each line will have a maximum of three random numbers with the last line having three or less depending on the quantity requested.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 12/16/1999 RANDOM NUMBER GENERATOR TIME: 14:22

AUDIT: EXAMPLE OF OUTPUT GOING TO THE PRINTER

SEED NUMBER: 42702.26 FRAME SIZE: 1,000

10 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER
10 -	50	9 -	321	3 -	359
8 -	389	2 -	484	4 -	490
1 -	491	7 -	569	5 -	733
6 -	815				

==========NEW PAGE==========

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 12/16/1999 RANDOM NUMBER GENERATOR TIME: 14:22

AUDIT: EXAMPLE OF OUTPUT GOING TO THE PRINTER

SEED NUMBER: 42702.26 FRAME SIZE: 1,000

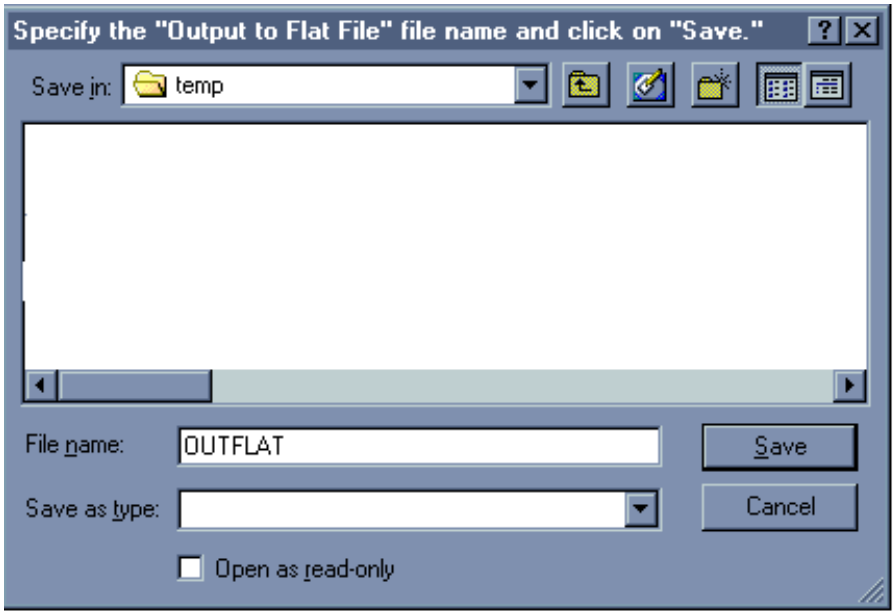
4 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER
11 -	599	12 -	648	13 -	921
14 -	592				

SUMMATION OF RANDOM NUMBERS 7,461

Output to a Flat File

The output following the screen insert is the text file created by selecting “Flat File” as one of the output options. Preceding this output is the standard Windows “Save” screen if the **OUTPUT TO Flat File** is selected. Fill in the name of the file in the **File name** box. The name of the file for this illustration is C:\TEMP\OUTFLAT. This file is the output file created when generating 10 random numbers between 1 and 1000 along with four spares. Notice that the order of selection and the random values contain leading zero values. This file is often useful as an input file for selecting random records using a mainframe computer.



File C:\TEMP\OUTFLAT

00100000000050 µ The 10 random values start here.
00090000000321
00030000000359
00080000000389
00020000000484
00040000000490
00010000000491
00070000000569
00050000000733
00060000000815
00110000000599 µ The 4 spare values start here.
00120000000648
00130000000921
00140000000592

Attribute Appraisals

OVERVIEW

UNRESTRICTED

STRATIFIED

The purpose of an attribute sample is to determine the number of items that meet a given set of criteria. Typically, in such a sample methodology, the reviewer will conclude after analyzing a sample item that the item does or does not meet the stated criteria. The criteria may be as simple as whether or not an approving signature is on a document or as complicated as to whether or not a tax return met all of the IRS rules and regulations. The conclusion, however, is usually stated as yes or no.

While the conclusion may be quite simple, the methods used to select the sample may require quite sophisticated appraisal techniques. The reviewer, for example, may want to group states by certain characteristics and then sample from each group of states and still be able to make an overall statement about a given set of criteria.

This package offers the user two appraisal methodologies in designing and performing a statistical sample. A brief example for using each program is given below. A detailed explanation of how to use each module is described later in this section.

Unrestricted

This module is typically used when an unrestricted sample has been drawn. A reviewer may have analyzed a sample of time cards from a pay period to determine the number of time cards that show employees arriving late for work. An unrestricted random sample of time cards was used to select the cards for review.

Stratified

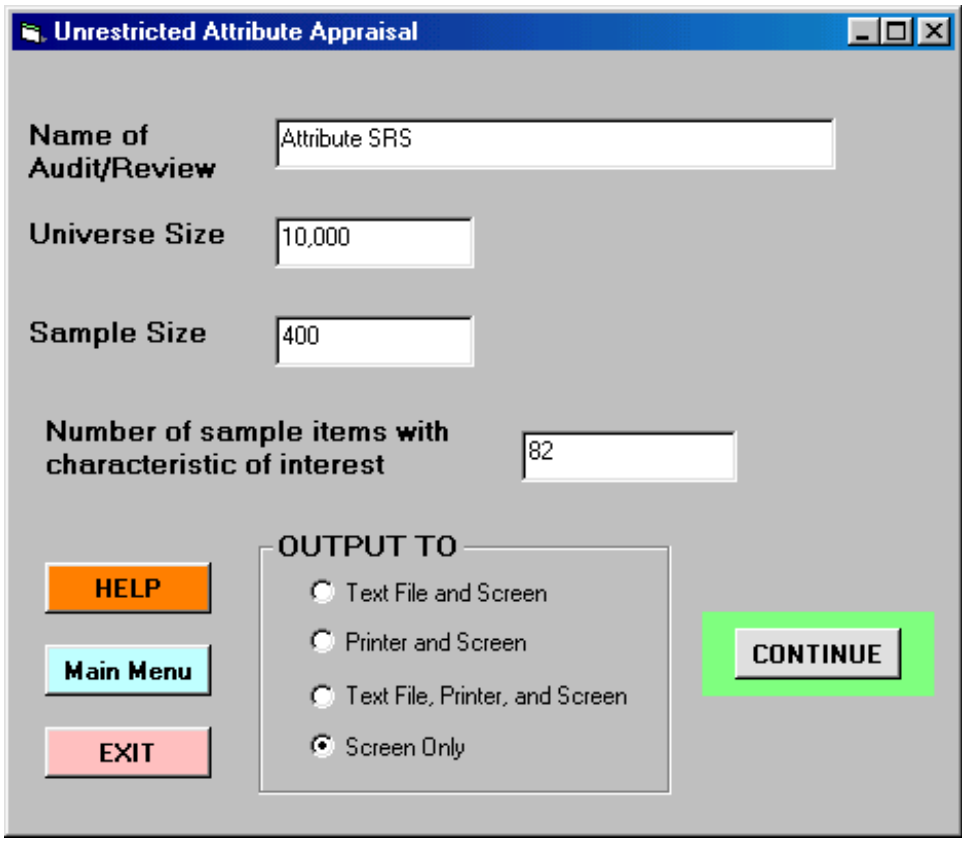
In certain cases the reviewer may want to divide the overall universe of transactions into two or more categories (strata). However, the reviewer may still want to be able to make a statistical statement about the overall universe. Expanding on the time card review mentioned above, the reviewer may want to place each time card into one of three categories: (1) clerical workers, (2) professional staff, or (3) management. A separate sample would be drawn from each group (stratum). This program will develop statistics for each stratum as well as providing overall statistics.

UNRESTRICTED

Purpose

This program performs an attribute appraisal on data input by the user based on an unrestricted random sample. Attribute sampling is used to determine how frequently an event or type of transaction occurs in a given universe. This type of sampling usually requires a yes or no (true or false) evaluation of each sampling unit by the user. The results are usually reported as a percentage.

Input Screen



NOTE: Example is for illustrative purposes only. The sample size may not conform with the organization's minimum sample size standards.

Name of Audit/Review

This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

Universe Size

The universe size is the total number of items from which the sampled items were selected. The number should be entered without commas but upon exiting this box, the commas will be inserted. The maximum allowable universe size is 2,147,483,647.

Sample Size

The sample size is the quantity of items selected by the user from the universe. This number also should be entered without commas.

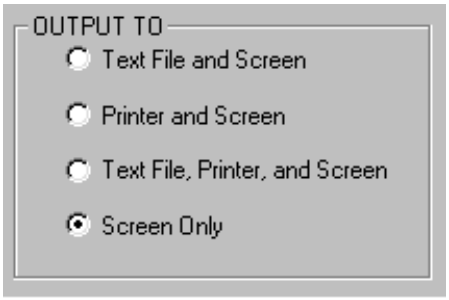
Number of Sample Items with the Characteristic of Interest

The user must establish evaluation criteria for all of the sample items. These criteria must be applied consistently to all items. The user needs to identify all sample items that have met the evaluation criteria ("characteristic of interest"). Depending on the purpose of the appraisal, the user would enter the number of items meeting the criteria or the number of items failing to meet the criteria. For example, if the user was looking at 100 documents to see if the documents had the proper approval signature, then the characteristic of interest would be the approval signature. If the evaluation of the sample showed that 88 documents out of 100 had the proper approval, the user would enter the response to the number of sample items with characteristic of interest as 88. The user could also enter 12 (i.e., 100 - 88) if the purpose of the appraisal was to estimate what percentage of the universe of documents did not have the approval signature. The program will also evaluate samples that have:

1. zero occurrences of the evaluation criteria
2. all sample items meeting the evaluation criteria.

Program Output

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by clicking the corresponding button.

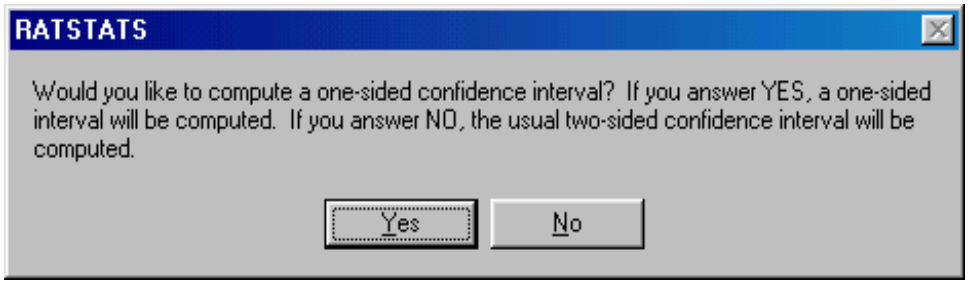


The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

The program will reprint the data supplied by the user (universe size, sample size and number of sample items with the characteristic of interest) and also provide the appraisal results. The following pieces of information will be displayed:

PROJECTED QUANTITY IN UNIVERSE	The proportion of sample items with the characteristic of interest multiplied by the universe size.
PERCENT	The proportion of sample items with the characteristic of interest displayed as a percentage.
CONFIDENCE LEVELS	The confidence (80%, 90%, or 95%) associated with the ability of the interval corresponding to contain the true proportion (or total number in the universe).
LOWER LIMIT	The lower boundary of the confidence interval. The limit is shown as both a number and percentage of the universe.
UPPER LIMIT	The upper boundary of the confidence interval. The limit is shown as both a number and percentage of the universe.

In the event the sample contains zero items having the characteristic of interest, the user will see the following screen:

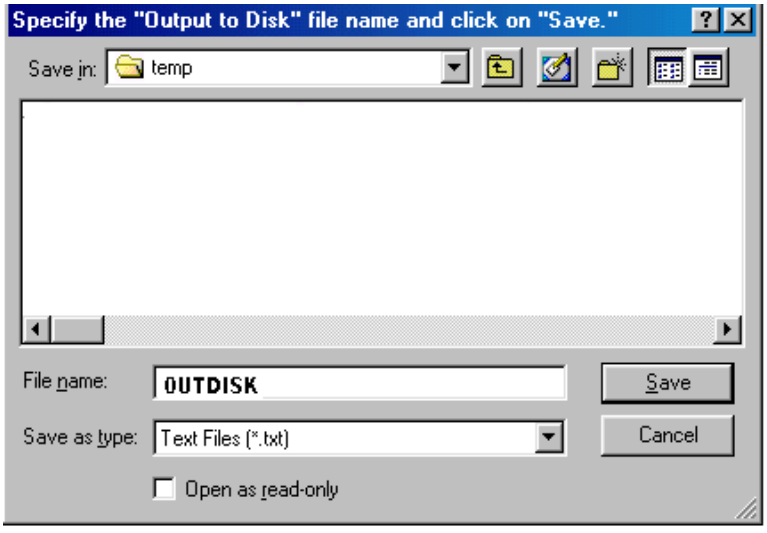


If the user responds with "Yes," the program will only compute the upper limit and the lower limit will not be computed. If the user responds with "No," the program will compute both the lower and upper limits.

In the event the number of sample items with the characteristic of interest is the same as the sample size, the user will also see the preceding screen. If the user responds with "Yes," the program will only compute the lower limit and the upper limit will not be computed. If the user responds with "No," the program will compute both the lower and upper limits.

Output to a Text File

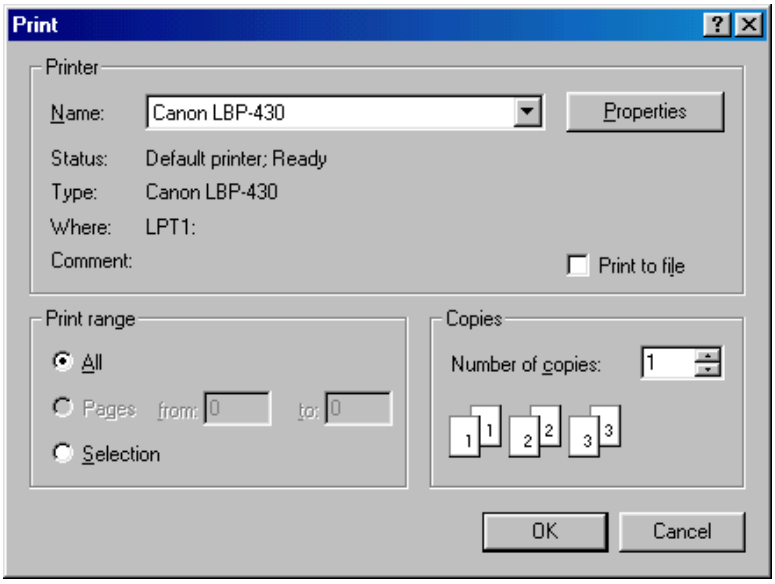
A text file will be saved with a ".txt" extension (e.g., C:\TEMP\OUTDISK.txt). Next, the user will see the standard Windows "Save" screen. Fill in the name of the file in the **File name** box. The file name for this illustration is C:\TEMP\OUTDISK.txt.



By clicking on the **Save** button, the user will return to the original input screen for this module.

Output to Printer

If the user selects the printer for output, the standard Windows "Print" dialog box (shown next) will appear. Select the appropriate printer and click on **OK**.



Format of Output: Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.txt, shown next. The printer output is identical.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 7/24/2001          Single Stage Attribute Appraisal          Time: 14:14
                           AUDIT/REVIEW: Attribute SRS


          UNIVERSE SIZE                                                10,000
          SAMPLE SIZE                                                  400
          CHARACTERISTIC(S) OF INTEREST
            QUANTITY IDENTIFIED IN SAMPLE                              82
            PROJECTED QUANTITY IN UNIVERSE                             2,050
            PERCENT                                                   20.500%
          STANDARD ERROR
            PROJECTED QUANTITY                                         198
            PERCENT                                                    1.978%
```

```
                                   CONFIDENCE LIMITS

                                 80% CONFIDENCE LEVEL
LOWER LIMIT - QUANTITY                          1,796
              PERCENT                          17.960%
UPPER LIMIT - QUANTITY                          2,326
              PERCENT                          23.260%

                                 90% CONFIDENCE LEVEL
LOWER LIMIT - QUANTITY                          1,729
              PERCENT                          17.290%
UPPER LIMIT - QUANTITY                          2,403
              PERCENT                          24.030%

                                 95% CONFIDENCE LEVEL
LOWER LIMIT - QUANTITY                          1,673
              PERCENT                          16.730%
UPPER LIMIT - QUANTITY                          2,470
              PERCENT                          24.700%
```

In the event that (1) the sample contained no items with the characteristic of interest or (2) the number of sample items with the characteristic of interest is equal to the sample size, the user has the option of selecting either a one-sided confidence interval (only one of the limits is determined) or the usual two-sided interval (both limits are determined). For example, the following output will be obtained for the situation where no items of interest were found in a sample of size 400 and the user answered "Yes" to the query "Would you like to compute a one-sided confidence interval? If you answer YES, a one-sided interval will be computed. If you answer NO, the usual two-sided confidence interval will be computed."

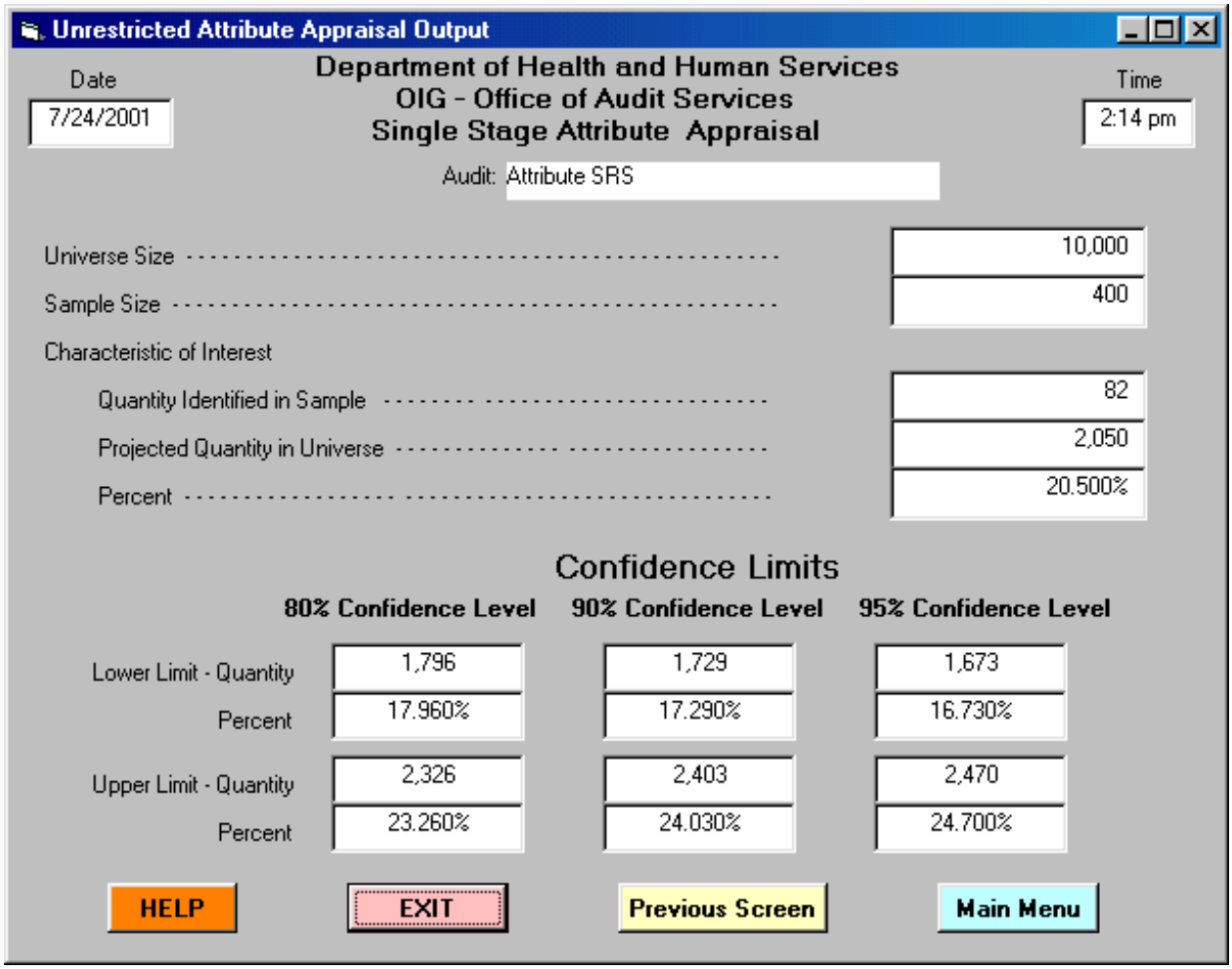
```
                  Department of Health and Human Services
                       OIG - Office of Audit Services
Date: 7/24/2001        Single Stage Attribute Appraisal            Time: 14:51
                          AUDIT/REVIEW: Attribute SRS


          UNIVERSE SIZE                                                10,000
          SAMPLE SIZE                                                  400
          CHARACTERISTIC(S) OF INTEREST
            QUANTITY IDENTIFIED IN SAMPLE                              0
            PROJECTED QUANTITY IN UNIVERSE                             0
            PERCENT                                                    .000%
          STANDARD ERROR
            PROJECTED QUANTITY                                         0
            PERCENT                                                    .000%

                                   CONFIDENCE LIMITS

                                 80% CONFIDENCE LEVEL
          UPPER LIMIT - QUANTITY                                         39
                        PERCENT                                        .390%
```

```
                                    90% CONFIDENCE LEVEL
      UPPER LIMIT - QUANTITY                          56
                    PERCENT                         .560%

                                    95% CONFIDENCE LEVEL
      UPPER LIMIT - QUANTITY                          73
                    PERCENT                         .730%

   SINCE NO ITEMS HAVING THE CHARACTERISTIC(S) OF INTEREST WERE FOUND IN THE
   SAMPLE, THE PROGRAM HAS CALCULATED ONLY THE MAXIMUM NUMBER OF ITEMS
   HAVING THE CHARACTERISTIC(S) OF INTEREST IN THE UNIVERSE.
```

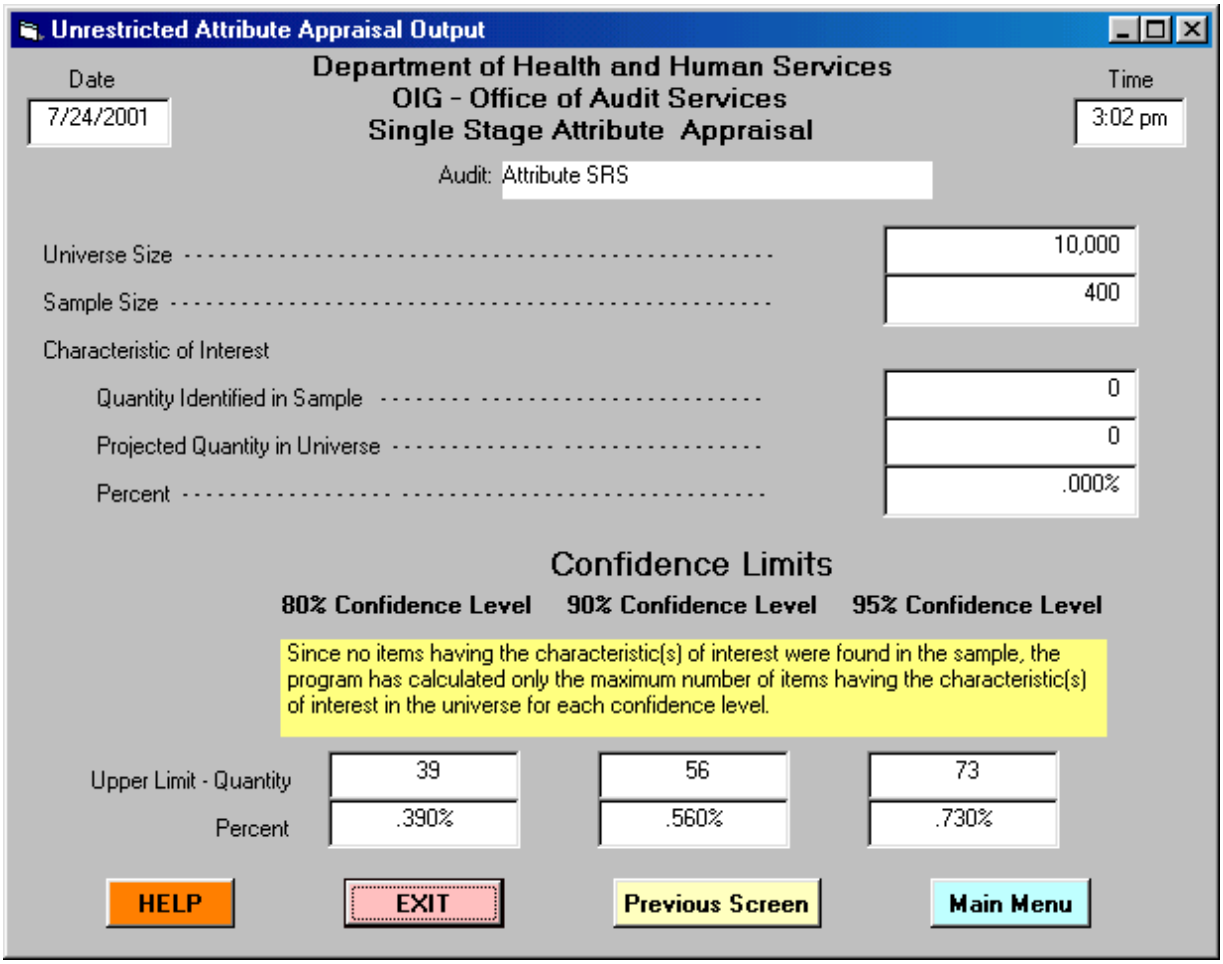
If the preceding sample had contained 400 items having the characteristic of interest and the user specified a one-sided interval, the above output would only contain the lower limits, along with the message "SINCE ALL SAMPLE ITEMS CONTAINED THE CHARACTERISTIC(S) OF INTEREST, THE PROGRAM HAS CALCULATED ONLY THE MINIMUM NUMBER OF ITEMS IN THE UNIVERSE HAVING THE CHARACTERISTIC(S) OF INTEREST."

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected is disk or printer. The following screen is the summary provided for this illustration.



The following screen will be obtained for the situation where no items of interest were found in a sample of size 400 and the user answered "Yes" to the query "Would you like to compute a one-sided confidence interval? If you answer YES, a one-sided interval will be computed. If you answer NO, the usual two-sided confidence interval will be computed."



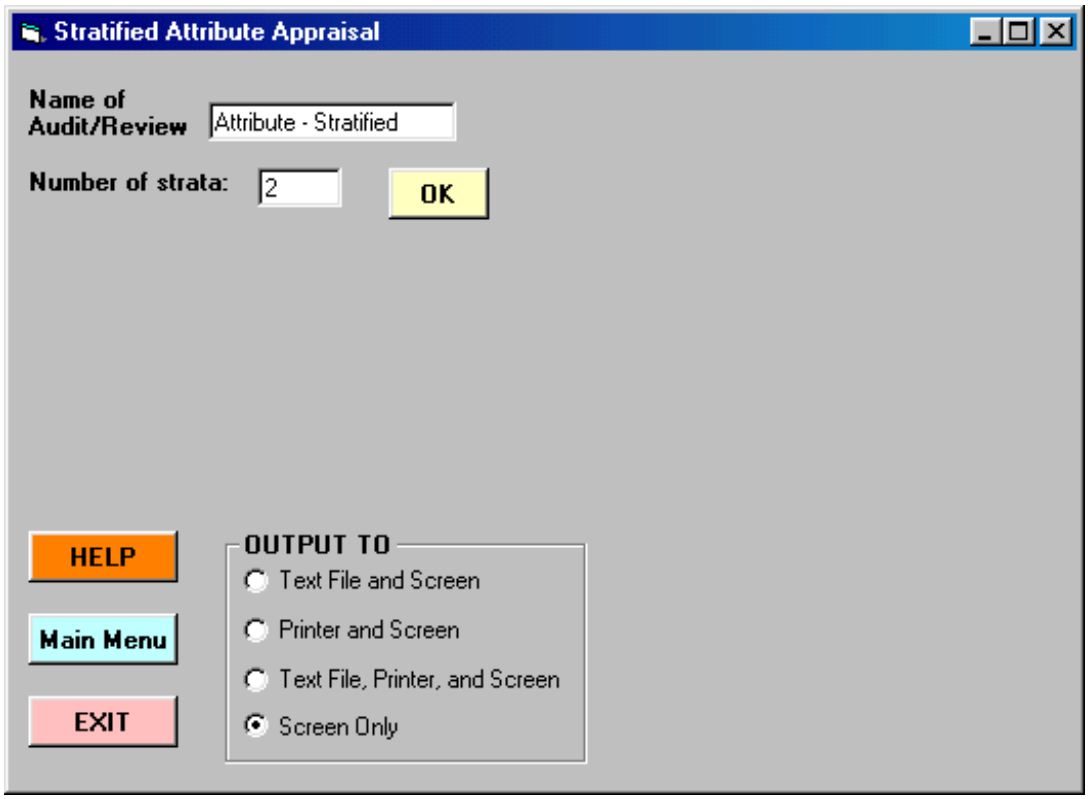
If the preceding sample had contained 400 items having the characteristic of interest and the user specified a one-sided interval, the preceding screen would only contain the lower limits, along with the message "Since all sample items had the characteristic(s) of interest, the program has calculated only the minimum number of items having the characteristic(s) of interest in the universe for each confidence level."

STRATIFIED

Purpose

This program provides a stratified attribute appraisal from the user's input to a series of prompts. Attribute sampling is used to determine how frequently an event or type of transaction occurs in a given universe. This type of sampling usually requires a yes or no (true or false) evaluation of each sampling unit by the user. The results are usually reported as a percentage.

Input Screen

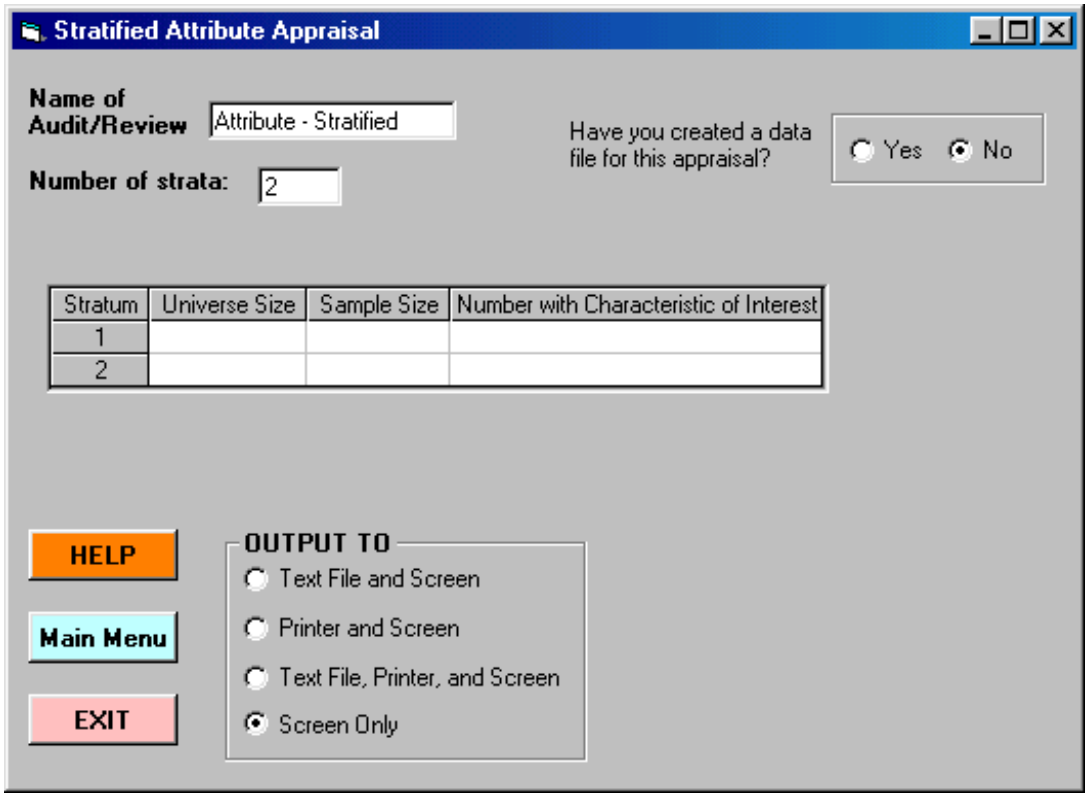


Name of Audit/Review

This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

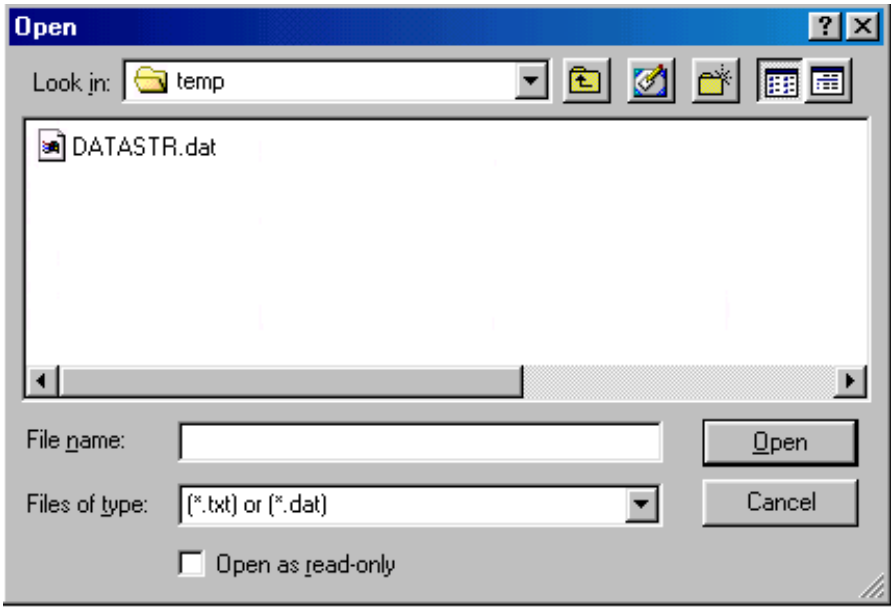
Number of Strata

After entering the name of the audit/review, the user must enter the number of strata. After entering the name and the number of strata, click on **OK**. The following screen will appear.



Have You Created A Data File?

The user may create a data file prior to running this appraisal. For each stratum the user needs to enter the universe size, sample size and number of items with the characteristic of interest. The program also allows the user to edit the values and save the modified file. The values may also be entered from the keyboard and subsequently saved as a data file. If the user has not created a data file, select "No" in the above box. If "Yes" is selected, the standard Windows "Open" file screen will appear. The window will contain all ".txt" and ".dat" files in the selected directory. To view all files in this directory, the user may change the file type to "All files."



Click on the file name (DATASTR.dat for this illustration) and then click on **Open** (or simply double click on file DATASTR.dat). If a file name is entered, the values obtained from the file will be displayed on the screen for review by the user. The format for the data should be as follows:

1000 100 2

Explanation:

1000 - This is the size of the universe of items from which a sample was drawn in a stratum.

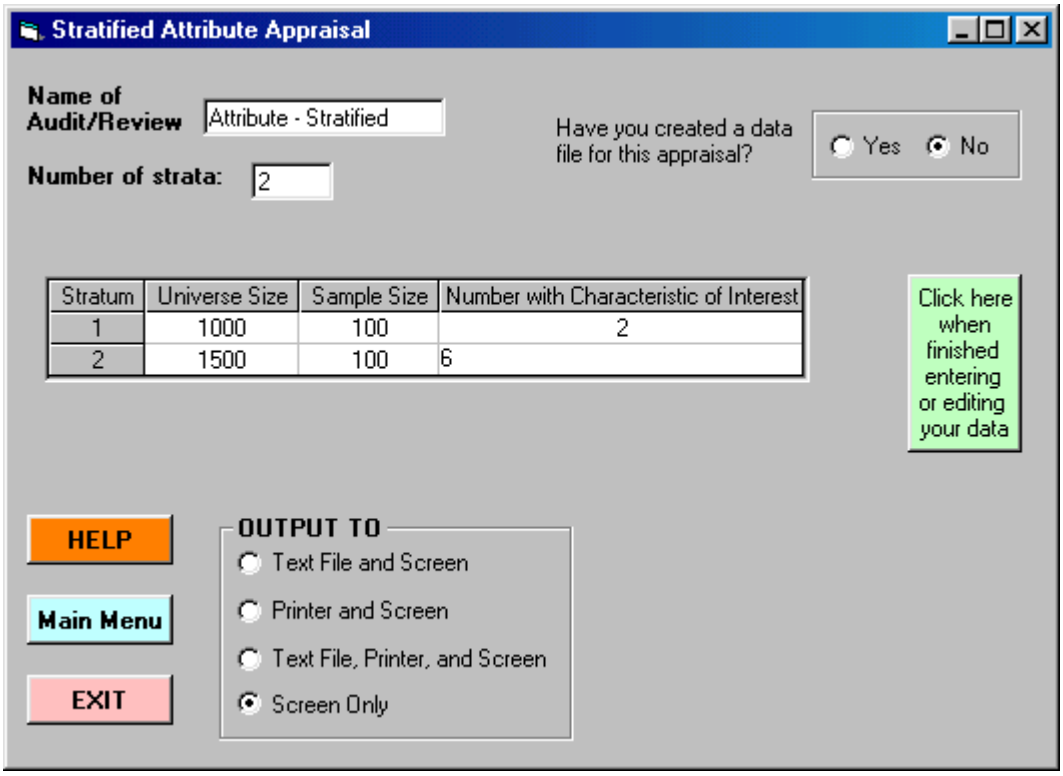
100 - This is the quantity of sample items that were reviewed.

2 - This is the quantity of items that met the criteria established for the sample review.

The program assumes a comma or a space as a delimiter between pieces of data. The user needs to be sure that in entering large numbers (e.g., 10000), commas are not used for readability (e.g., 10,000) since the program will assume that each comma is separating two pieces of data (e.g., 10 and 000). The user must not use dollar signs ($) or any other symbols in conjunction with the data as the program will assign a value of zero to any input that contains nonnumeric data.

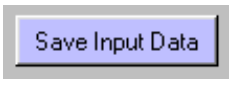
Entering the Sample Results from the Screen

The sample results can be entered from the screen form if a data file was not used to enter the results. For this input option, the user will be presented a grid as shown next. To illustrate, suppose stratum 1 contains 1,000 values, 100 of which are sampled, and 2 of the items contain the characteristic of interest. Also, stratum 2 contains 1,500 values, 100 of which are sampled, and 6 of the items contain the characteristic of interest.



NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

When all the values have been entered, click on **Click here when finished entering or editing your data**. **NOTE**: This grid structure does not allow the user to use the tab key to move from cell to cell. Each cell must be clicked on before entering its value.

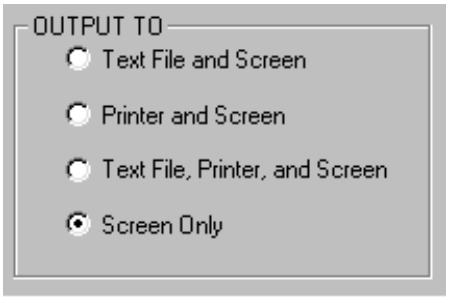


To save this data set, click on **Save Input Data**, which will appear after clicking on **Click here when finished entering or editing your data**. The standard Windows "Save As" screen will appear. Type the output file name alongside the **File name** box and click on **Save**.

To print this data set, click on **Print Input Data**, which will appear after clicking on **Click here when finished entering or editing your data**. Select a printer and click on **OK**.

Program Output

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by clicking the corresponding button. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.



The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The program always concludes with a summary on the screen.

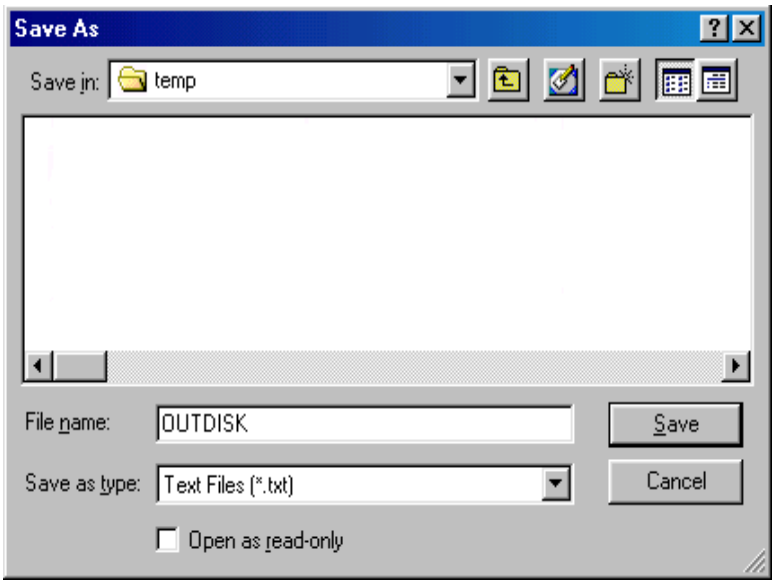
The program will output the data supplied by the user (sample size, number of items with characteristic of interest and universe size) for each stratum. In addition, the following pieces of information will be displayed:

RATIO	The percentage of sample items in each stratum with the characteristic of interest.
PROJ. ITEMS IN UNIVERSE	The result of applying the stratum ratio to the stratum universe. This is also calculated for the total universe.
PRECISION	The confidence interval half-width expressed as a percentage. Precision is calculated for each stratum and the universe as a whole.
LOWER LIMIT	The lower boundary of the confidence interval. The limit is shown as both a number and a percentage of the universe. The confidence levels are 80%, 90%, and 95%.

UPPER LIMIT The upper boundary of the confidence interval. The limit is shown as both a number and a percentage of the universe. The confidence levels are 80%, 90%, and 95%.

Output to a Text File

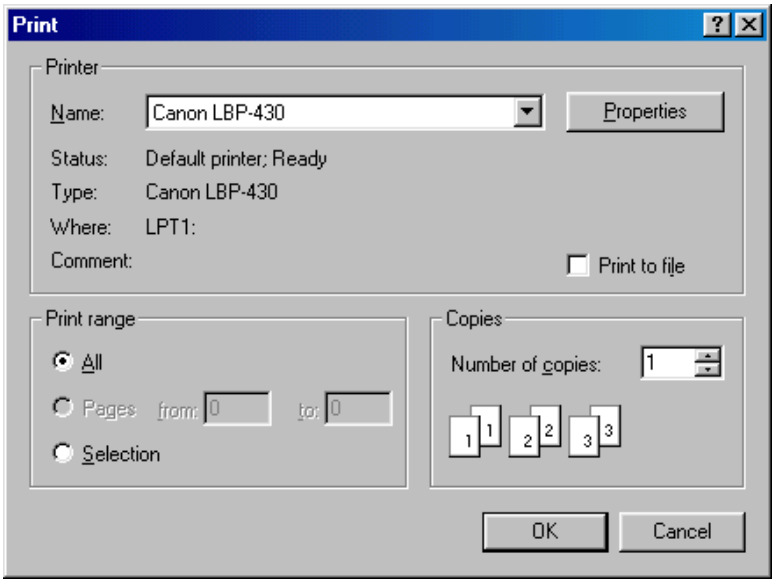
A text file will be saved with a ".txt" extension (e.g., C:\TEMP\OUTDISK.txt). Next, the user will see the standard Windows "Save As" screen. Fill in the name of the file in the **File name** box. The file name for this illustration is C:\TEMP\OUTDISK.txt.



By clicking on the **Save** button, the user will return to the original input screen for this module.

Output to Printer

If the user selects the printer for output, the standard Windows "Print" dialog box (shown next) will appear. Select the appropriate printer and click on **OK**.



Format of Output: Text File or Printer

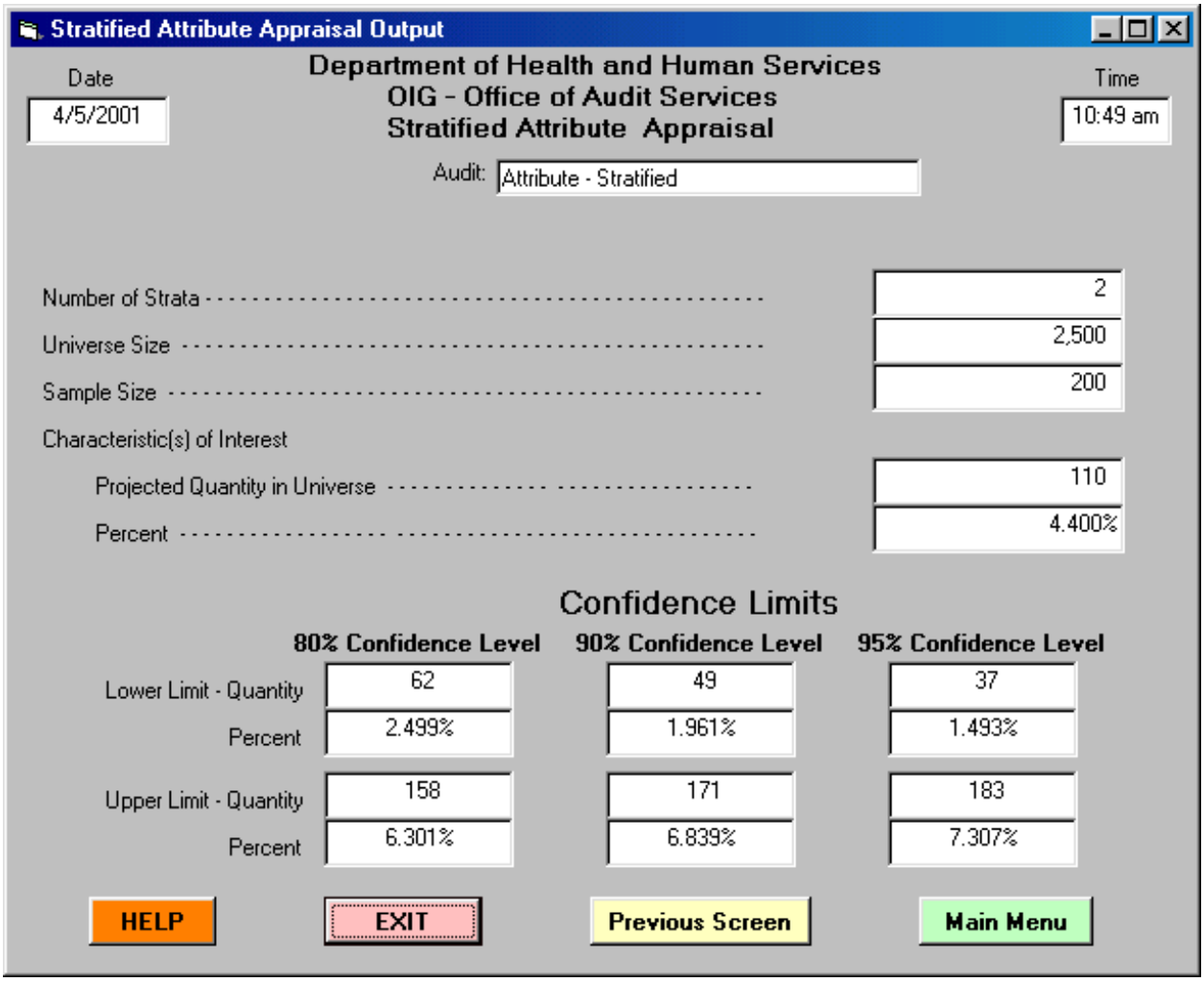
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.txt, shown next. The printer output is identical.

```
                DEPARTMENT OF HEALTH & HUMAN SERVICES
                   OIG - OFFICE OF AUDIT SERVICES
Date: 4/5/2001   STRATIFIED ATTRIBUTE APPRAISAL       Time: 10:55
                AUDIT/REVIEW: Attribute - Stratified

                                                          PROJ. ITEMS
STRATUM     SAMPLE   *ITEMS**   **RATIO*   *UNIVERSE*   IN UNIVERSE
=======     ======   ========   ========   ==========   ===========

      1        100          2     2.000%       1,000            20
      2        100          6     6.000%       1,500            90

COMBINED       200          8     4.400%       2,500           110

STANDARD ERROR:                   1.483%          37
```

STRATUM	PRECISION AT 80% CL	PRECISION AT 90% CL	PRECISION AT 95% CL
1	1.711%	2.196%	2.616%
2	2.955%	3.793%	4.519%
COMBINED	1.901%	2.439%	2.907%
LOWER LIMIT - QUANTITY	62	49	37
PERCENT	2.499%	1.961%	1.493%
UPPER LIMIT - QUANTITY	158	171	183
PERCENT	6.301%	6.839%	7.307%

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected is a text file or printer. The following screen is the summary provided for this illustration.



Variable Appraisals

OVERVIEW

UNRESTRICTED

STRATIFIED

The purpose of a variable sample is to determine a quantitative characteristic or set of characteristics about a population. The reviewer may want to determine the dollar value of an inventory or the amount of duplicate payments made by an organization. These types of estimates can be made with a variable sample.

This package offers the user two appraisal methodologies when designing and performing a variable statistical sample. A brief example of when to use each module is given below. A detailed explanation of how to use each module is described later in this section.

Unrestricted

This module is used when an unrestricted sample has been drawn. A reviewer may want to determine the value of an organization's inventory. The reviewer may have drawn an unrestricted random sample of inventory cards and has analyzed the related inventory to determine the actual value. Based on this sample, the reviewer could estimate the total inventory of the organization.

Stratified

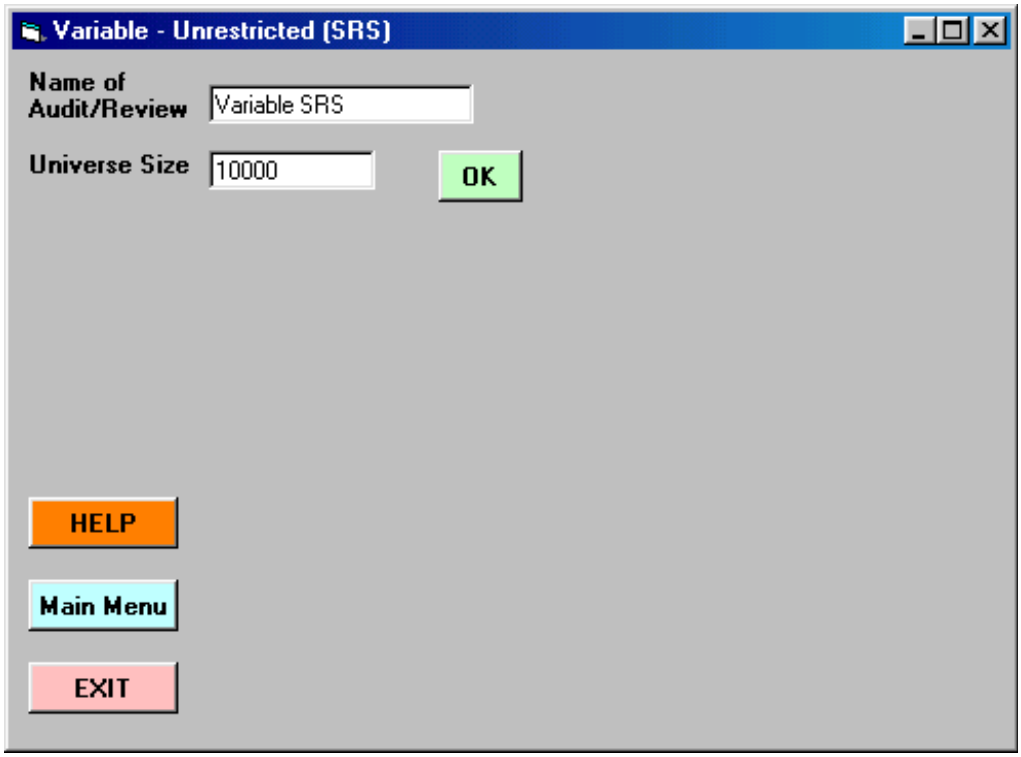
The reviewer may wish to spend more resources analyzing certain inventory items as compared to other items. One approach to accomplish this objective would be to stratify the inventory items into two or more categories (strata). One category (stratum) could be for more sensitive items (e.g., high value items). The appraisal can give the reviewer an estimate for each category as well as an overall approximation of the inventory.

UNRESTRICTED

Purpose

This program performs a variable appraisal on a data file previously created by the user based on information gathered from an unrestricted random sample. Variable sampling is used to estimate quantitative characteristics. For each sampling unit the user obtains one or more numeric pieces of information about an event or item. The user has the option of obtaining and appraising from one numeric piece of information per sample item (e.g., Examined amount) to as many as three pieces of information per sample item (i.e., Examined, Audited and Difference amounts). If the user decides to appraise all three pieces of information, only two of the pieces of data may be entered and the third will be calculated by the program. The variable appraisal program assumes that some variation exists between values. If no variation exists, then there is no need to run this appraisal program.

Input Screen



Name of Audit/Review

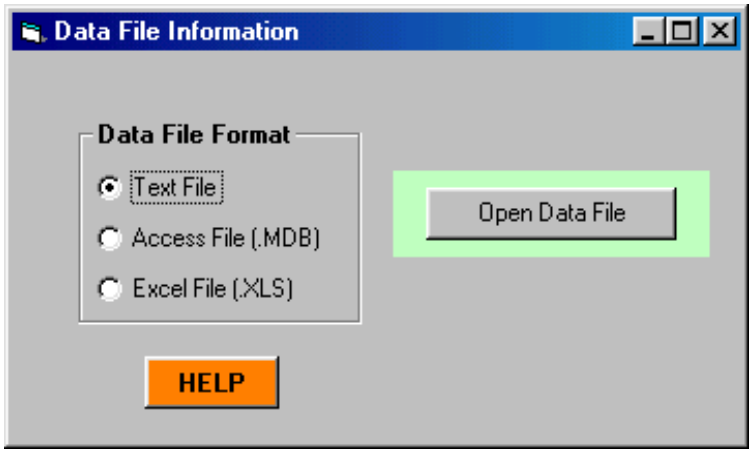
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

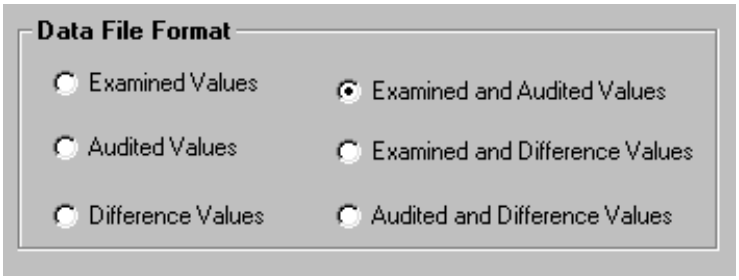
Universe Size

The universe size is the total number of items from which the sampled items were selected. The number should be entered without commas but upon exiting this box, the commas will be inserted. This number will be used in estimating universe parameters.

Format of Input File

After entering the above information, click on **OK**. The following screen will appear. The input file format can be a text file, a table within an Access database, or an Excel spreadsheet. After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open" file screen, locate the input file and double click on it. The program will then return to the original screen.





Prior to executing this program, the user must create a data file that contains certain identifying data and one or two pieces of information for each sample unit selected. Each data line consists of a line number for that sampling unit followed by the first piece of information (a numeric value) the user wants to appraise (i.e., examined, audited, or difference value). If two or more pieces of information will be appraised and the examined amount is one of the values, then the examined amount must be the first piece of data entered for each sampling unit. If only the audited and difference amounts are being appraised, then the audited amount must be the first piece of data entered. The second piece of information may be the numeric difference between the examined value and the amount accepted by the user or the audited amount if the examined amount was the first piece of data entered.

Regardless of the software used to create the data file, the format should be as follows:

7483 289.99 43.00

Explanation:

7483 - This is a number assigned by the user. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.

289.99 - This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.

43.00 - If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of $289.99 claimed by a vendor, only $246.99 was actually owed, then the difference amount entered would be $43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least

one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. **Note** : If there are two pieces of information for each sampling unit (e.g., examined value and audited value), each line in the data file must contain only the sample item number and the corresponding two values (as illustrated in data set DATASRS below). Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs ($). The program assumes one or more spaces as the only delimiter between pieces of data.

The user has the option at the end of the file of adding the following line:

9999 3E33

This line serves as a sentinel. The line number may be any number selected by the user. In stratified samples, the program will use this line to indicate where data from one stratum ends. This line is optional in an unrestricted variable appraisal data file.

Input from a Text File

The data file containing the above information may be stored in a text file format. This file need not have an extension, such as .TXT. There are several ways the user may create this file. The easiest approach would be with a word processing package (e.g., Word) or a print file created with a spreadsheet package (e.g., Excel).

The first 20 rows and last three rows of data set DATASRS are shown below.

```
1   300   267
2   900   774
3   300   255
4   200   174
5   900   810
6   700   560
7   1000  820
8   100   80
9   900   765
10  700   630
11  700   630
12  400   332
13  300   255
14  100   84
15  200   168
16  100   88
17  600   528
```

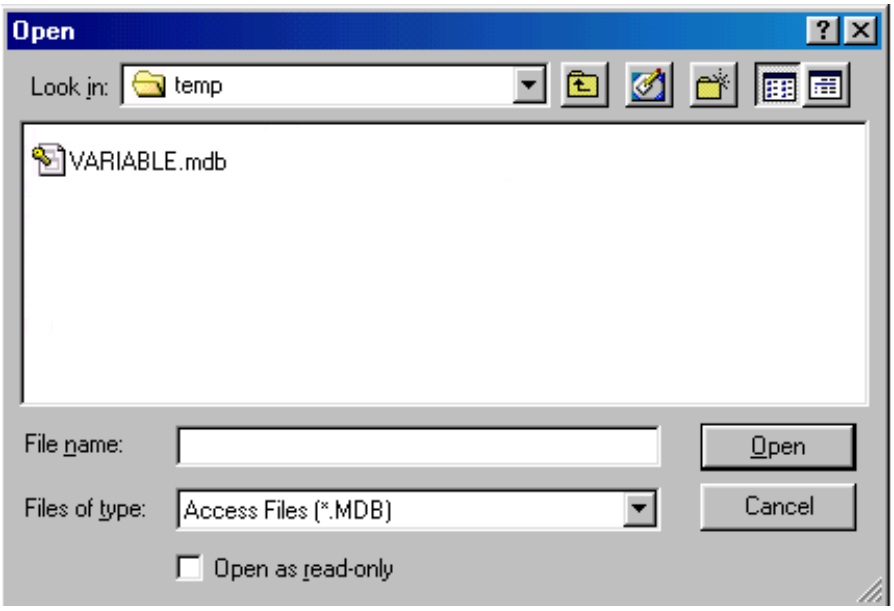
Data set DATASRS

```
18   400    340
19   900    747
20   1000   800
  .
  .
  .
48    300   237
49    500   435
50    100   86
```

NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

Input from an Access Table

The input file may be stored in a table within an Access database. Select the name of the database containing the input table in the preceding **Open Data File** step. This database must have the standard Access extension (.MDB). The name of the database for this illustration is C:\TEMP\VARIABLE.mdb. The user may double click on the VARIABLE database name or single click on it and the **Open** button.



The name of the Access table within database VARIABLE.mdb for this illustration is DATASRS. The following table shows the first 20 rows of Access table DATASRS. The field name for the first column ("Line-Number" in the illustration) is arbitrary and is not used by the program at any point.

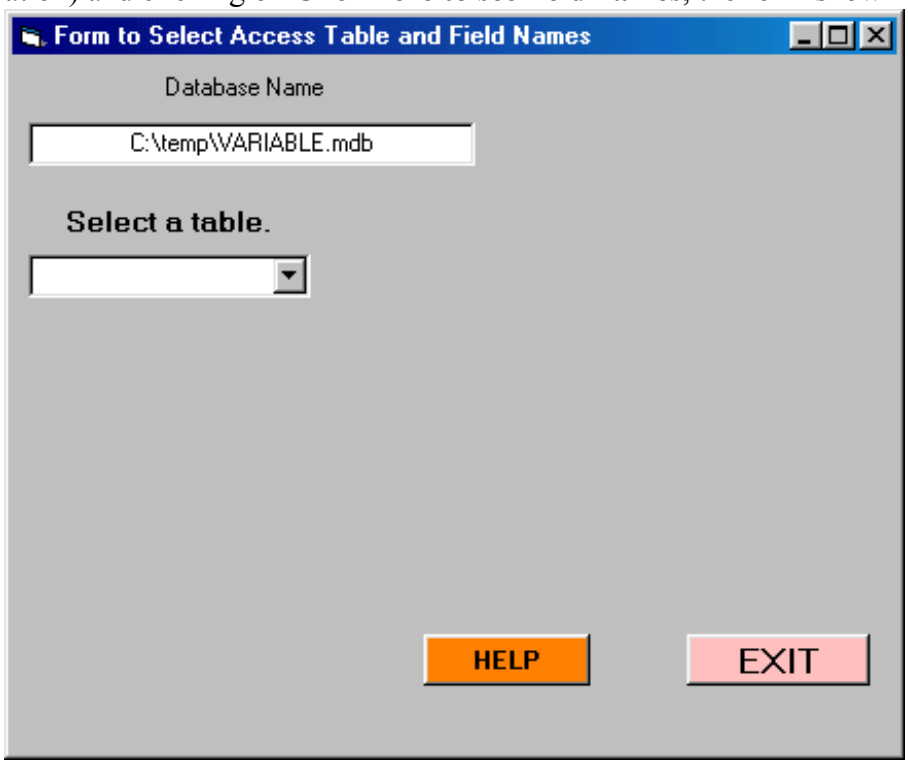
DATASRS : Table

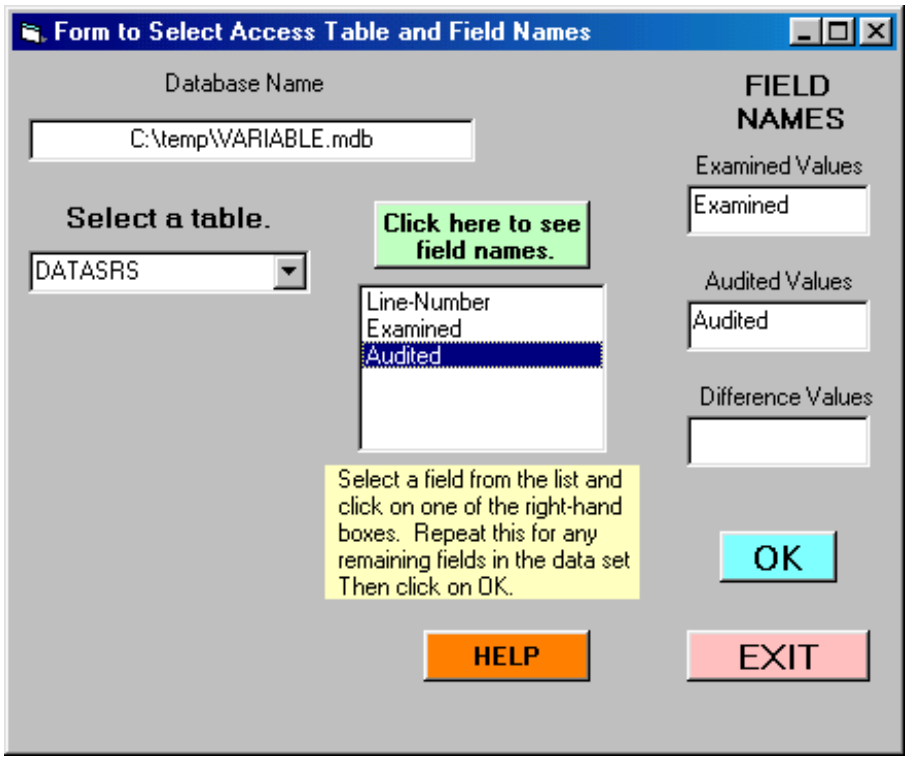
Line-Number	Examined	Audited
1	300	267
2	900	774
3	300	255
4	200	174
5	900	810
6	700	560
7	1000	820
8	100	80
9	900	765
10	700	630
11	700	630
12	400	332
13	300	255
14	100	84
15	200	168
16	100	88
17	600	528
18	400	340
19	900	747
20	1000	800

NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

In the Data File Information screen, after clicking on the **OK** button, the user will be asked to select the name of the table within the selected database using the following form. Click on the down arrow under **Select a table**. After selecting a table from the drop-down list (DATASRS

for this illustration) and clicking on **Click here to see field names**, the form shown next will appear.

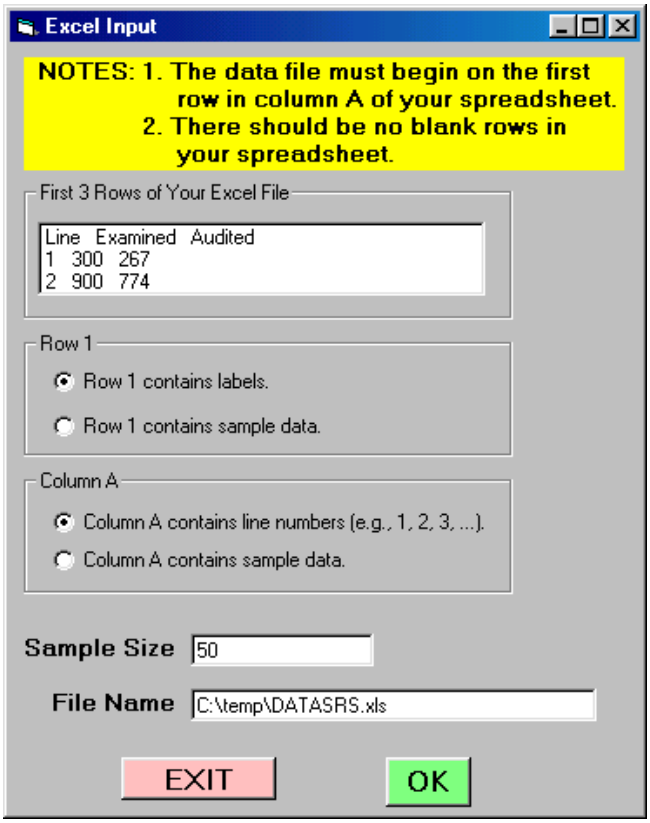




To select the field names, click on the field name for the field containing the first piece of information in the input file ("Examined" in this illustration) and then click on the top right-hand box. The field name will then appear in this box. Repeat this for the field name of the second field in this table ("Audited") and click on the middle right-hand box to specify this field name. When the field names have been specified, click on **OK**. The program will then continue processing. **NOTE**: When the user returns to the original input screen, the data file format (Examined and Audited Values for this illustration) will be selected based on responses within the preceding Form to Select Access Table and Field Names. When using an Access input file, the user cannot change this data file format option after returning to the original input screen.

Input from an Excel Spreadsheet

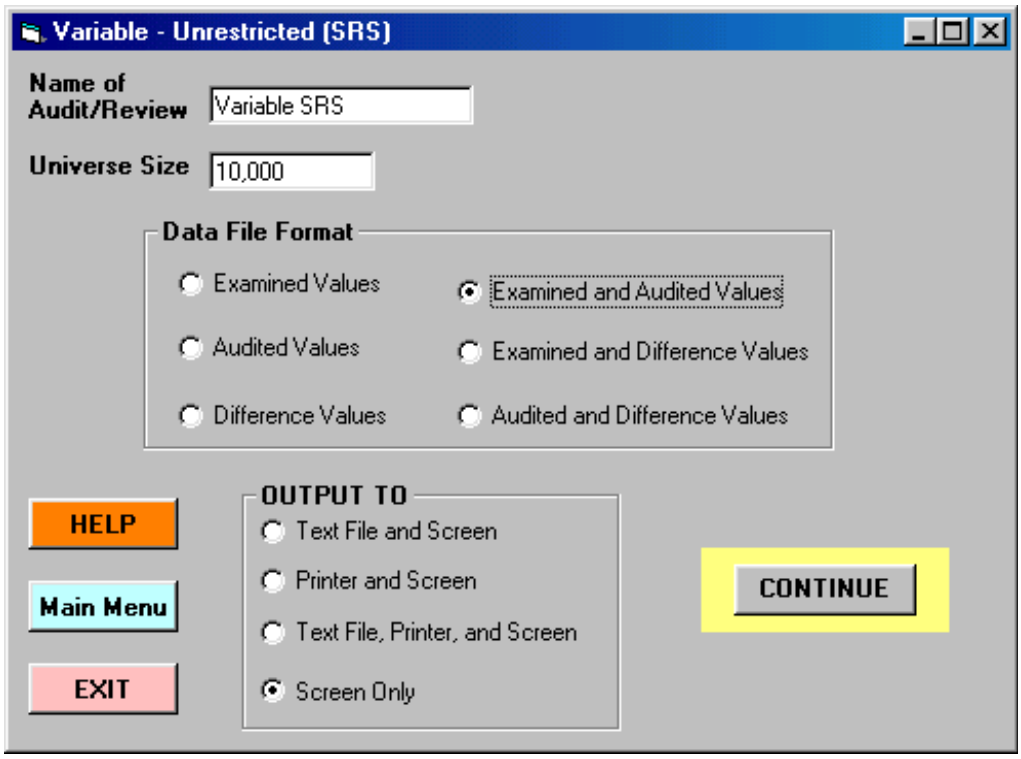
The input file can be stored in an Excel spreadsheet. Select the name of the spreadsheet containing the input table in the preceding Open Data File step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATASRS.XLS will be used. The first 20 rows of this file are shown immediately following the Excel Input screen. This particular file contains labels (variable names) in the first row and line numbers (1, 2, 3, ...) in column A. The corresponding options were selected in the Excel Input screen. The line numbers are optional. Had column A contained the examined values, the second option in the Column A frame in the Excel Input screen should have been selected.



NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

	A	B	C
1	**Line**	**Examined**	**Audited**
2	1	300	267
3	2	900	774
4	3	300	255
5	4	200	174
6	5	900	810
7	6	700	560
8	7	1000	820
9	8	100	80
10	9	900	765
11	10	700	630
12	11	700	630
13	12	400	332
14	13	300	255
15	14	100	84
16	15	200	168
17	16	100	88
18	17	600	528
19	18	400	340
20	19	900	747

After the data file has been opened, the user will see the complete input screen.



Output Options

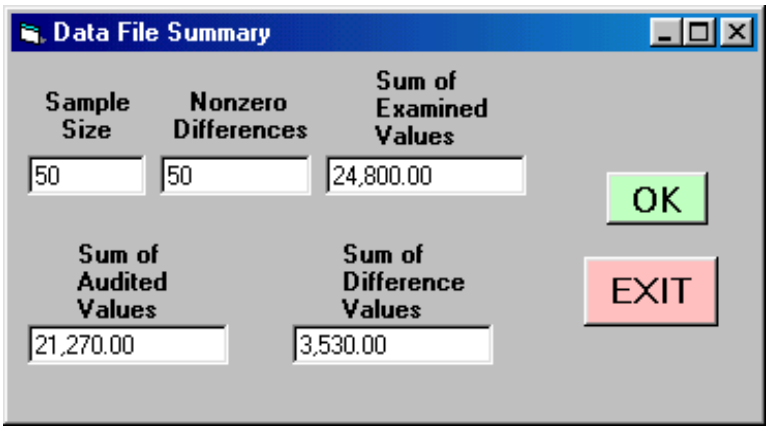
The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the program will prompt for a file name. The standard Windows "Save" file screen will appear. The user should type in the file name in the designated box and click on **Save**. The output file will be saved with a ".txt" extension.

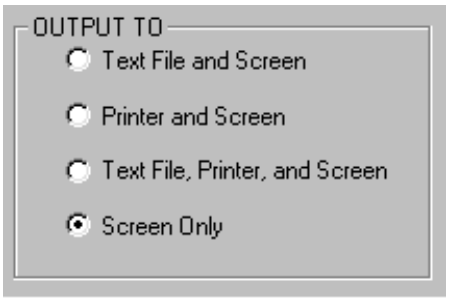
By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. The user can then retrieve the file through a word processing package, such as Word.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the printer to use for the program output.

After selecting the output format, click on **CONTINUE**. The next screen to appear is the data file summary, shown next. At this point the user should reconcile the values to determine that the data file is complete and accurate. Click on **OK** to continue or **EXIT** to exit the program.



Program Output



The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed:

MEAN -	The average value for the sample items appraised. It is obtained by summing the items in the sample and dividing the result by the number of items in the sample.
UNIVERSE -	This is the quantity of the items from which the sample was drawn. The results of the sample will be projected to the universe using this value.
STANDARD DEVIATION -	A measurement of the variation of the sample items about the average value (mean).
STANDARD ERROR -	A measurement of the variation of the sample mean with respect to all possible means for this universe and sample size.

SKEWNESS -	A measure of the symmetry of the frequency distribution of the sample items. Accounting universes are usually right skewed (majority of items have a low value while a few items have a high value).
KURTOSIS -	A measure of the peakedness or flatness of the frequency distribution of the sample values.
POINT ESTIMATE -	A single estimate for the universe total based on the sample mean multiplied by the universe size.
CONFIDENCE LEVEL -	The confidence (80%, 90%, 95%) associated with the ability of the corresponding interval to contain the true mean (or universe total).
LOWER LIMIT -	The lower bound of the confidence interval derived by subtracting the precision amount from the point estimate.
UPPER LIMIT -	The upper bound of the confidence interval derived by adding the precision amount to the point estimate.
PRECISION AMOUNT -	A measurement of the closeness of the sample estimate of the universe total and the corresponding unknown universe value. The precision amount is calculated by multiplying the standard error by the universe size and multiplying the result by the appropriate factor ("t" value) corresponding to the desired confidence level.
PRECISION PERCENT -	This is the result of dividing the precision amount by the point estimate and stating the result as a percentage.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.txt, shown next. The printer output is identical.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
Date: 4/5/2001 VARIABLE UNRESTRICTED APPRAISAL Time: 11:23
AUDIT/REVIEW: Variable SRS
DATA FILE USED: C:\temp\DATASRS.txt

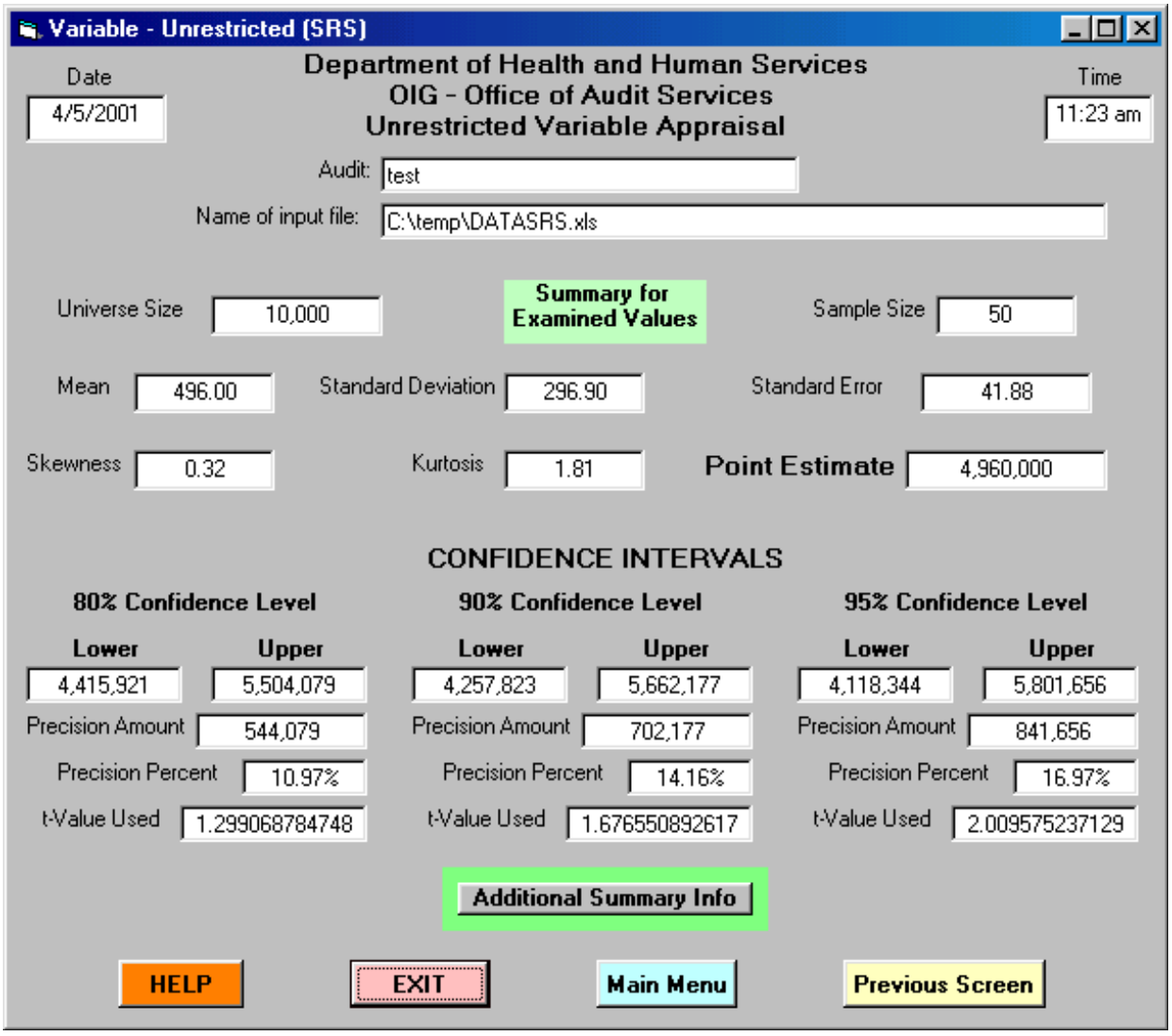
SAMPLE SIZE	EXAMINED VALUE	NONZERO DIFFS	TOTAL OF DIFF VALUES	TOTAL OF AUD VALUES
50	24,800.00	50	3,530.00	21,270.00

---------- E X A M I N E D ----------		
MEAN / UNIVERSE	496.00	10,000
STANDARD DEVIATION	296.90	
STANDARD ERROR	41.88	
SKEWNESS	.32	
KURTOSIS	1.81	
POINT ESTIMATE	4,960,000	
	CONFIDENCE LIMITS	
	80% CONFIDENCE LEVEL	
LOWER LIMIT	4,415,921	
UPPER LIMIT	5,504,079	
PRECISION AMOUNT	544,079	
PRECISION PERCENT	10.97%	
T-VALUE USED	1.299068784748	
	90% CONFIDENCE LEVEL	
LOWER LIMIT	4,257,823	
UPPER LIMIT	5,662,177	
PRECISION AMOUNT	702,177	
PRECISION PERCENT	14.16%	
T-VALUE USED	1.676550892617	
	95% CONFIDENCE LEVEL	
LOWER LIMIT	4,118,344	
UPPER LIMIT	5,801,656	
PRECISION AMOUNT	841,656	
PRECISION PERCENT	16.97%	
T-VALUE USED	2.009575237129	
---------- A U D I T E D ----------		
MEAN / UNIVERSE	425.40	10,000
STANDARD DEVIATION	256.20	
STANDARD ERROR	36.14	
SKEWNESS	.30	
KURTOSIS	1.78	
POINT ESTIMATE	4,254,000	

```
                              CONFIDENCE LIMITS
                             80% CONFIDENCE LEVEL
LOWER LIMIT                             3,784,500
UPPER LIMIT                             4,723,500
PRECISION AMOUNT                          469,500
PRECISION PERCENT                           11.04%
T-VALUE USED                       1.299068784748

                             90% CONFIDENCE LEVEL
LOWER LIMIT                             3,648,074
UPPER LIMIT                             4,859,926
PRECISION AMOUNT                          605,926
PRECISION PERCENT                           14.24%
T-VALUE USED                       1.676550892617

                             95% CONFIDENCE LEVEL
LOWER LIMIT                             3,527,715
UPPER LIMIT                             4,980,285
PRECISION AMOUNT                          726,285
PRECISION PERCENT                           17.07%
T-VALUE USED                       2.009575237129

---------------------- D I F F E R E N C E ----------------------
MEAN / UNIVERSE                            70.60             10,000
STANDARD DEVIATION                         48.25
STANDARD ERROR                              6.81
SKEWNESS                                     .64
KURTOSIS                                    2.98
POINT ESTIMATE                           706,000

                              CONFIDENCE LIMITS
                             80% CONFIDENCE LEVEL
LOWER LIMIT                               617,575
UPPER LIMIT                               794,425
PRECISION AMOUNT                           88,425
PRECISION PERCENT                           12.52%
T-VALUE USED                       1.299068784748

                             90% CONFIDENCE LEVEL
LOWER LIMIT                               591,881
UPPER LIMIT                               820,119
PRECISION AMOUNT                          114,119
PRECISION PERCENT                           16.16%
T-VALUE USED                       1.676550892617

                             95% CONFIDENCE LEVEL
LOWER LIMIT                               569,213
UPPER LIMIT                               842,787
PRECISION AMOUNT                          136,787
PRECISION PERCENT                           19.37%
T-VALUE USED                       2.009575237129
```

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected includes a text file or printer. The following screen is the summary of the examined values for this illustration. To see the summary for the audited values, click on **Additional Summary Info** at the bottom of this form. To see the summary for the difference values, click again on **Additional Summary Info**.



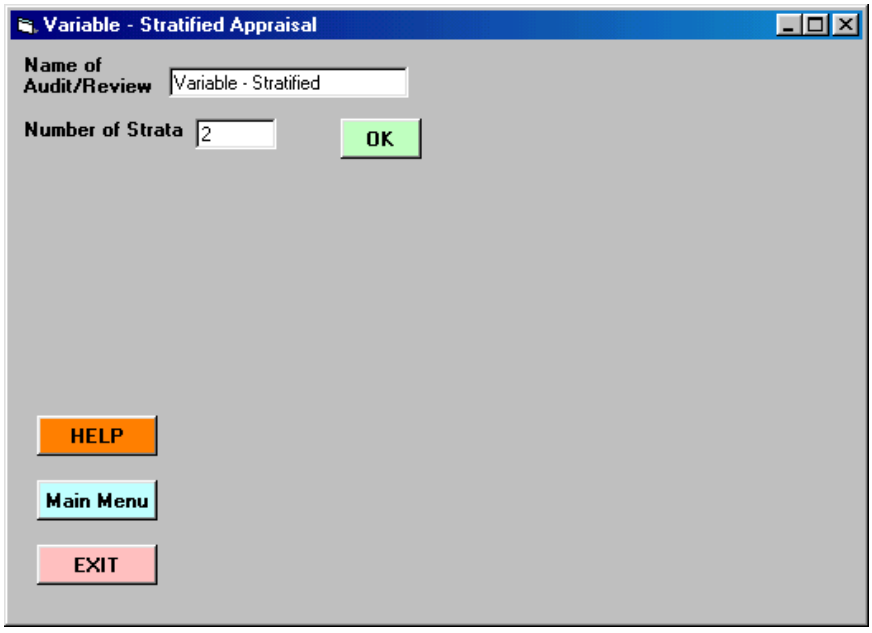
STRATIFIED

Purpose

This program performs a stratified variable appraisal on a data file previously created by the user based on information gathered from a stratified random sample. Variable sampling is used to estimate quantitative characteristics. For each sampling unit the user obtains one or more numeric pieces of information about an event or item. The user has the option of obtaining and appraising from one numeric piece of information per sample item (e.g., Examined amount) to as many as three pieces of information per sample item (i.e., Examined, Audited and Difference amounts). If the user decides to appraise all three pieces of information, only two of the pieces of data may be entered and the third will be calculated by the program. The variable appraisal program assumes that some variation exists between values.

The user normally selects stratification to improve sample efficiency. The area of interest, for example, may be placed into segments (strata) based on value of items (e.g., high and low dollar value of transactions) or the sensitivity of items reviewed (e.g., entertainment and payroll costs). The program allows for a maximum of 50 strata to be appraised. The user must know the universe size of each stratum in order to use this methodology.

Input Screen

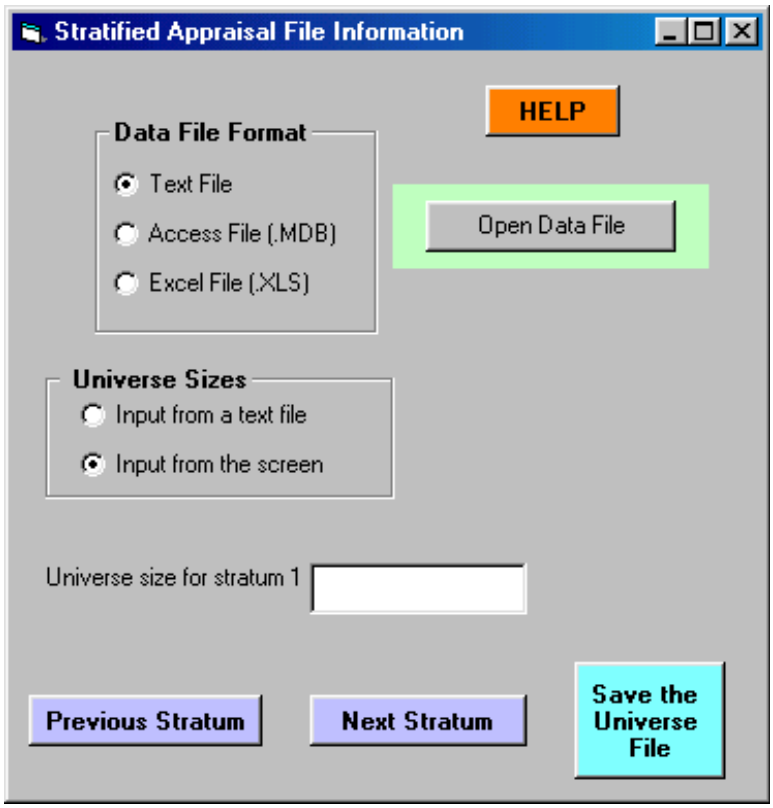


Name of Audit/Review

This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

Number of Strata

After entering the name of the audit/review, the user must enter the number of strata. The maximum number of strata is 50. After entering the name and the number of strata, click on the **OK** button and the following screen will appear. The **Universe Sizes** frame (and everything below it) will be invisible until the data file is opened.

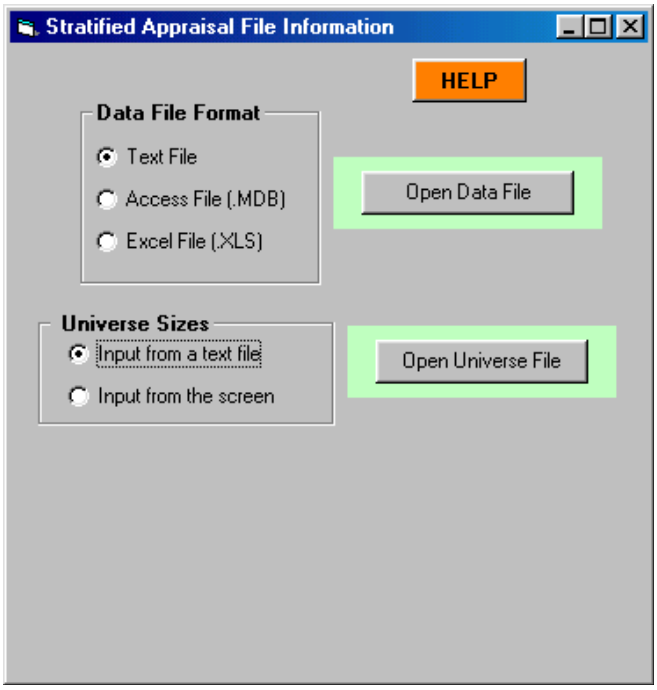


Opening the Data File

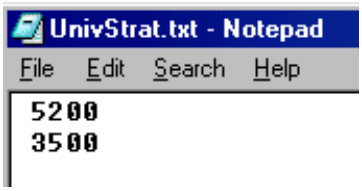
The input file format can be a text file, a table within an Access database, or an Excel spreadsheet. After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open" file screen, locate the input file and double click on it.

Universe Sizes

The user has the option of creating a file of universe sizes prior to executing the program or interactively entering the values during program execution. The format for this file is simply the universe sizes, in order by stratum, separated by one or more spaces. Commas are allowed in this data file. If the user selects "Input from a text file" in the Universe Sizes list, then the **Open Universe File** button will appear in the preceding form as shown next. To open this file, click on this button and use the standard Windows "Open" file screen to locate the universe file (e.g., UnivStrat.txt - - shown next). After this file has been opened, the **Read Universe File** button will appear. After clicking on this button, the remaining portion of the Stratified Appraisal File Information screen will be visible (including a button labeled **OK**) and will contain the universe size for the first primary unit (see "Entering the Universe Sizes from the Screen" on the next page). Click on the **OK** button to return to the original screen.



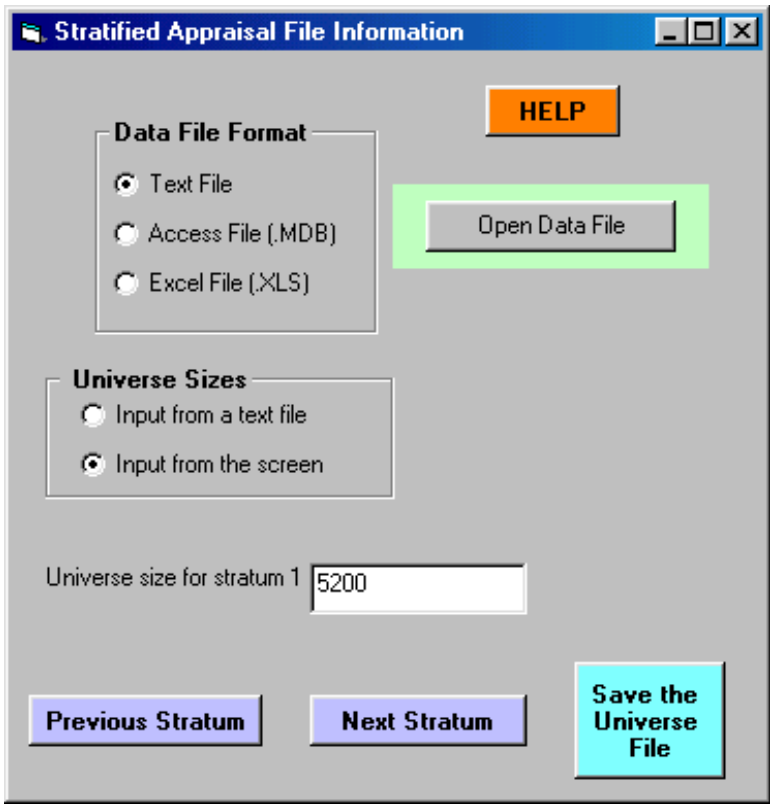
Stratified Files Information Screen



File UnivStrat.txt

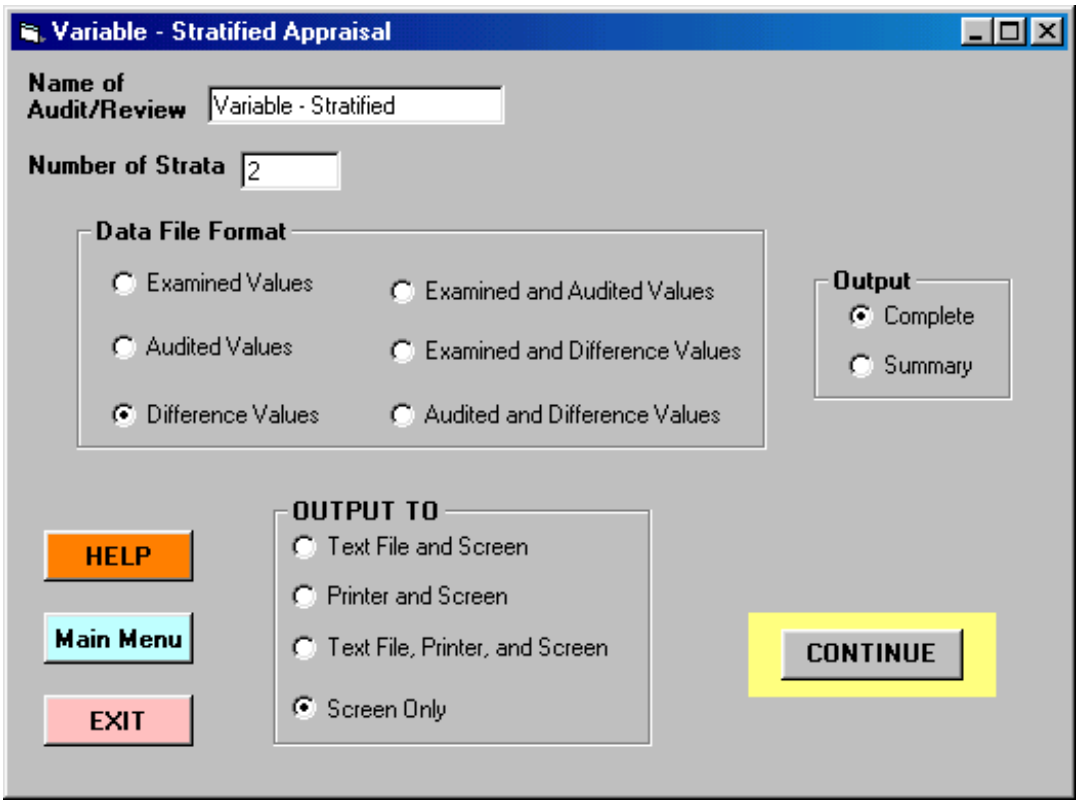
Entering the Universe Sizes from the Screen

If the user selects "Input from the screen" in the Universe Sizes list, the following screen will appear. Enter the universe size for stratum 1 (5200 for this illustration). To enter the size for stratum 2, click on **Next Stratum**.

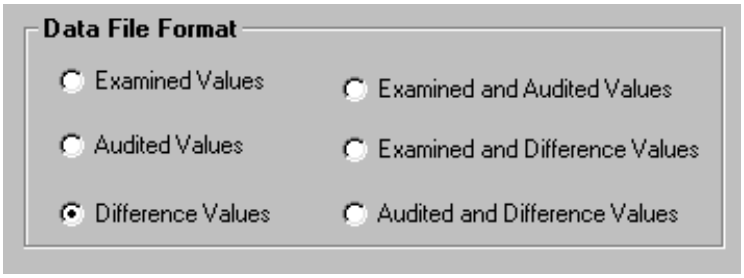


Whether or not the user entered the universe sizes from a text file, the user may view/edit the universe sizes by clicking on **Next Stratum** and **Previous Stratum** and making the desired changes. To save the universe file, click on **Save the Universe File**. The standard Windows "Save" file screen will appear. Enter a file name in the file name box (or select a file if replacing one) and click on the **Save** button.

When both files have been opened (or the data file has been opened and the universe sizes specified on the screen), click on the **OK** button and the screen shown next will appear.



Format of Input File



Prior to executing this program, the user must create a data file that contains certain identifying data and one or two pieces of information for each sample unit selected within each stratum. Each data line consists of a line number for that sampling unit followed by the first piece of information (a numeric value) the user wants to appraise (i.e., Examined, Audited, or Difference value). If two or more pieces of information will be appraised and the examined amount is one of the values, then the examined amount must be the first piece of data entered for each sampling unit. If only the audited and difference amounts are being appraised, then the audited amount must be the first piece of data entered. The second piece of information may be the numeric

difference between the examined value and the amount accepted by the user or the audited amount if the examined amount was the first piece of data entered.

Regardless of the software used to create the data file, the format should be as follows:

7483 289.99 43.00

Explanation:

7483 - This is a number assigned by the user. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.

289.99 - This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.

43.00 - If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of $289.99 claimed by a vendor, only $246.99 was actually owed, then the difference amount entered would be $43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. **Note** : If there are two pieces of information for each sampling unit (e.g., examined value and audited value), each line in the data file must contain only the sample item number and the corresponding two values. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs ($). The program assumes one or more spaces as the only delimiter between pieces of data.

The user will separate strata in the data file by entering the following line at the end of each stratum of values:

9999 3E33

This line serves as a sentinel and is optional after the last stratum in the file. The line number may be any number selected by the user.

Input from a Text File

The data file containing the above information may be stored in a text file format. There are several ways the user may create this file. The easiest approach would be with a word processing package (e.g., Word) or a print file created with a spreadsheet package (e.g., Excel).

A portion of data set DATASTRAT.txt is shown below. The data file contains 25 observations in each stratum. The rows shown after rows 25 and 50 contain the end-of-stratum indicators.

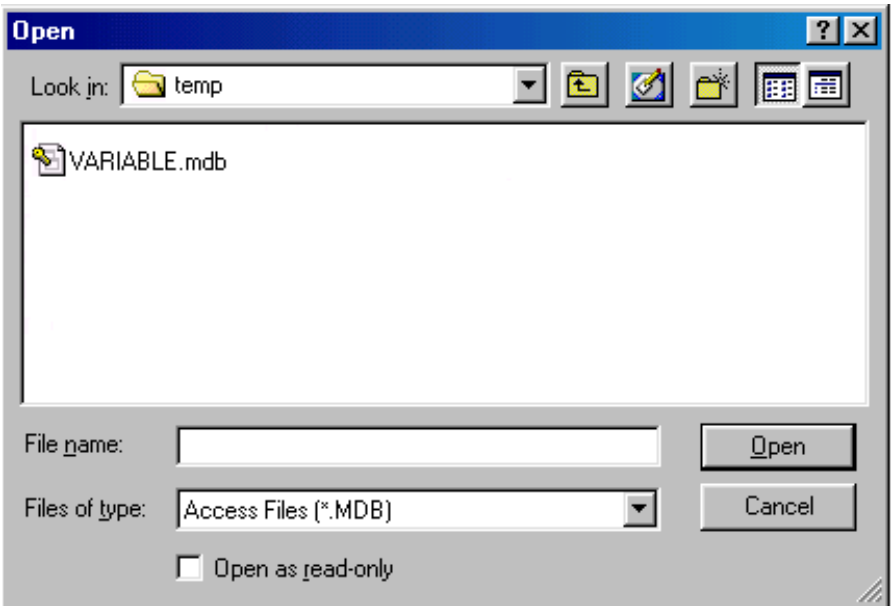
```
 1  80
 2  43
 3 133
 4 125
 5 116
 .
 .
 .
21 127
22 105
23 102
24  69
25  76
9999 3E33
26 354
27 328
28 313
29 250
30 261
 .
 .
 .
46 295
47 277
48 355
49 314
50 277
9999 3E33
```

Data file DATASTRAT.txt

NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

Input from an Access Table

The input file may be stored in a table within an Access database. Select the name of the database containing the input table in the preceding Open Data File step. This database must have the standard Access extension (.MDB). The name of the database for this illustration is C:\TEMP\VARIABLE.mdb. The user may double click on the VARIABLE database name or single click on it and the **Open** button.



NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

The name of the Access table for this illustration is DATASTRAT. The following table shows the first five rows. The field name for the first column ("Line-Number" in the illustration) is arbitrary and is not used by the program at any point.

DATASTRAT : Table

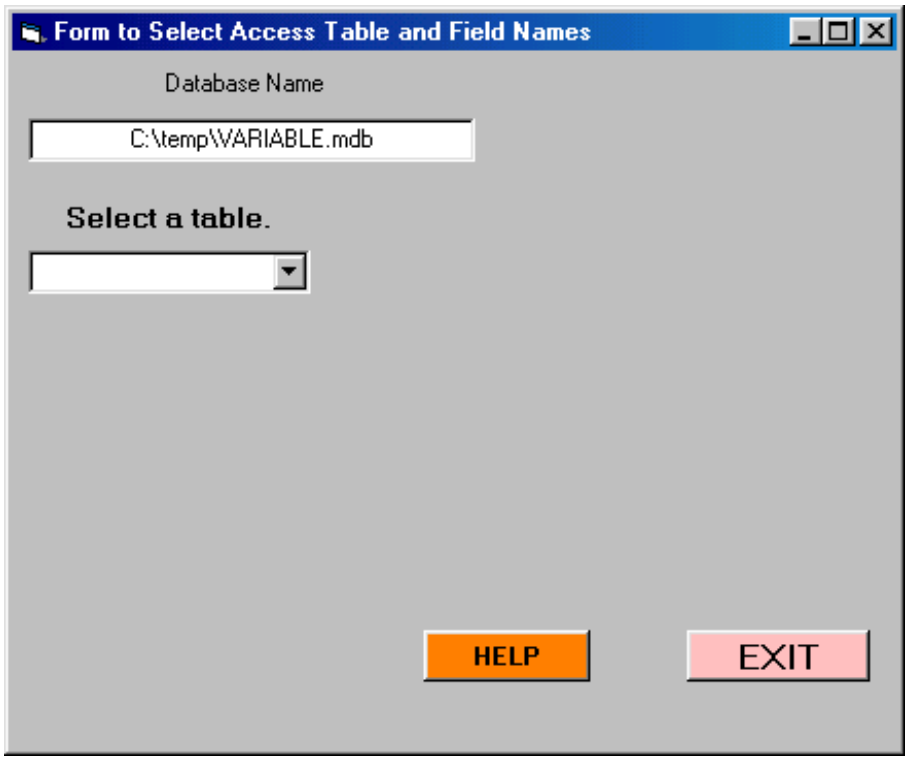
Line-Number	Difference
1	80
2	43
3	133
4	125
5	116

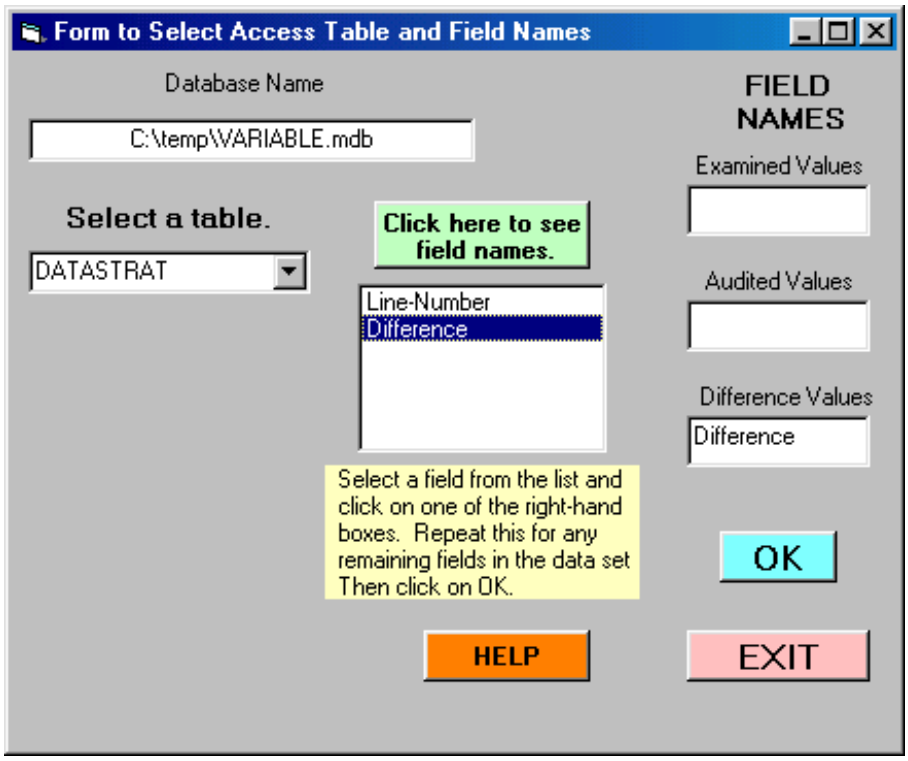
The data file contains 25 observation in each stratum. The last three rows of each stratum and the end-of-stratum indicators are shown below.

23	102
24	69
25	76
9999	3E+33
26	354

48	355
49	314
50	277
9999	3E+33

After entering the universe sizes and clicking on the **OK** button, the user will be asked to select the name of the table within the selected database using the following form. Click on the down arrow under **Select a table**. After selecting a table from the drop-down list (DATASTRAT for this illustration) and clicking on **Click here to see field names**, the form shown next will appear.

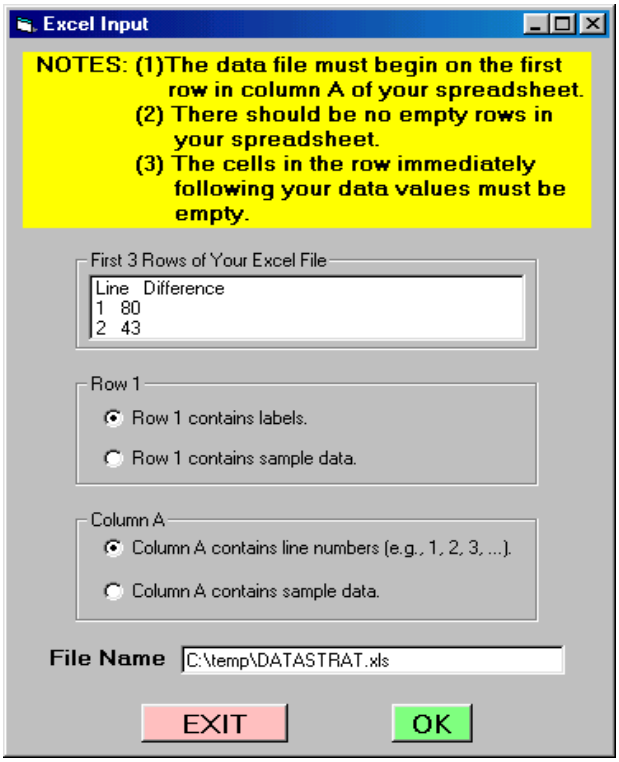

Form to Select Access Table and Field Names
Database Name
C:\temp\VARIABLE.mdb
Select a table.
HELP
EXIT


Form to Select Access Table and Field Names
Database Name
C:\temp\VARIABLE.mdb
Select a table.
DATASTRAT
Click here to see field names.
Line-Number
Difference
Select a field from the list and click on one of the right-hand boxes. Repeat this for any remaining fields in the data set Then click on OK.
FIELD NAMES
Examined Values
Audited Values
Difference Values
Difference
OK
HELP
EXIT

To select the field names, click on the field name for the field containing the first piece of information in the input file ("Difference" in this illustration) and click on the box labeled "Difference Values." The field name will appear in this box. Repeat this procedure for any remaining field names in this table (there are none for this illustration). When all the field names have been specified, click on **OK**. The program will continue processing. **NOTE**: When the user returns to the original input screen, the data file format (Difference Values for this illustration) will be selected, based on responses within the preceding Form to Select Access Table and Field Names. When using an Access input file, the user cannot change this data file format option after returning to the original input screen.

Input from an Excel Spreadsheet

The input file can be stored in an Excel spreadsheet. Select the name of the spreadsheet containing the input table in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATASTRAT.XLS will be used. The first five rows of this file are shown immediately following the Excel Input screen. The data file contains 25 observation in each stratum. Rows 25 through 27 and the last four lines in the second stratum containing the end-of-stratum indicators are also shown. This particular file contains labels (variable names) in the first row and line numbers (1, 2, 3, ...) in column A. The corresponding options were selected in the Excel Input screen. The line numbers are optional. Had column A contained the difference values, the second option in the Column A frame in the Excel Input screen should have been selected.



NOTE: Example is for illustrative purposes only. The sample sizes may not conform with the organization's minimum sample size standards.

	A	B
1	Line	Difference
2	1	80
3	2	43
4	3	133
5	4	125
6	5	116

25	24	69
26	25	76
27	9999	3.00E+33

49	47	277
50	48	355
51	49	314
52	50	277
53	9999	3.00E+33

Complete or Summary Output

The user may want to reduce printed output by having only the summary of the appraisal created. The default is for the complete appraisal output.

Output Options

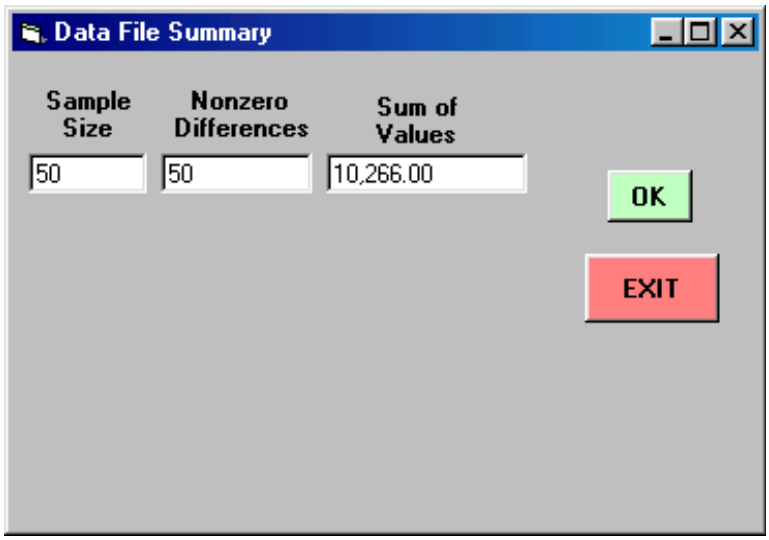
The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the program will prompt for a file name. The standard Windows "Open" file screen will appear. The user should type in the file name in the designated box and click on **Save**. The output file will be saved with a ".txt" extension.

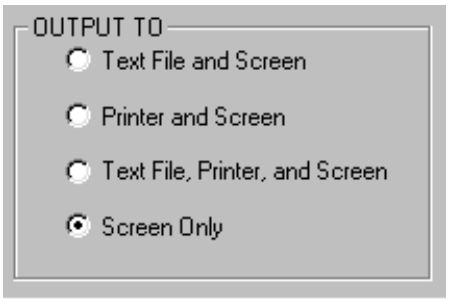
By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. The user can then retrieve the file through a word processing package, such as Word.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the printer to use for the program output.

After selecting the output format, click on **CONTINUE**. The next thing to appear is the data file summary, shown next. At this point the user should reconcile the values to determine that the data file is complete and accurate. Click on **OK** to continue or **EXIT** to exit the program.



Program Output



The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed:

MEAN - The average value for the sample items appraised. It is obtained by summing the items in the sample and dividing the result by the number of items in the sample.

UNIVERSE - This is the quantity of the items from which the sample was drawn. The results of the sample will be projected to the universe using this value.

STANDARD DEVIATION- A measurement of the variation of the sample items about the average value (mean).

STRATUM STANDARD ERROR - A measurement of the variation of the sample mean with respect to all possible means for this stratum universe and this sample size.

OVERALL STANDARD ERROR - A measurement of the variation of the point estimate of the total with respect to all possible totals for this universe and these sample sizes.

SKEWNESS - A measure of the symmetry of the frequency distribution of the sample items. Accounting universes are usually right skewed (majority of items have a low value while a few items have a high value).

KURTOSIS - A measure of the peakedness or flatness of the frequency distribution of the sample values.

POINT ESTIMATE - A single estimate for the universe total based on the sample mean multiplied by the universe size.

CONFIDENCE LEVEL - The confidence (80%, 90%, 95%) associated with the ability of the corresponding interval to contain the true mean (or universe total).

LOWER LIMIT - The lower bound of the confidence interval derived by subtracting the precision amount from the point estimate.

UPPER LIMIT - The upper bound of the confidence interval derived by adding the precision amount to the point estimate.

STRATUM PRECISION AMOUNT - A measurement of the closeness of the sample estimate of the universe total and the corresponding unknown universe value. The precision amount is calculated by multiplying the stratum standard error by the stratum universe size and multiplying the result by the appropriate factor ("t" value) corresponding to the desired confidence level. For the examined (reviewed) appraisal, the stratum total may be known and should be reviewed by the user to see if, in fact, the actual value does fall within the confidence interval.

OVERALL PRECISION AMOUNT - A measurement of the closeness of the sample estimate of the universe total and the corresponding unknown universe value. The precision amount is calculated by multiplying the overall standard error by the appropriate factor ("Z" value) corresponding to the desired confidence level. For the examined (reviewed) appraisal, the universe total may be known and should be reviewed by the user to see if, in fact, the actual value does fall within the confidence interval.

PRECISION PERCENT - This is the result of dividing the precision amount by the point estimate and stating the result as a percentage.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTSTRAT.txt, shown next. The printer output is identical.

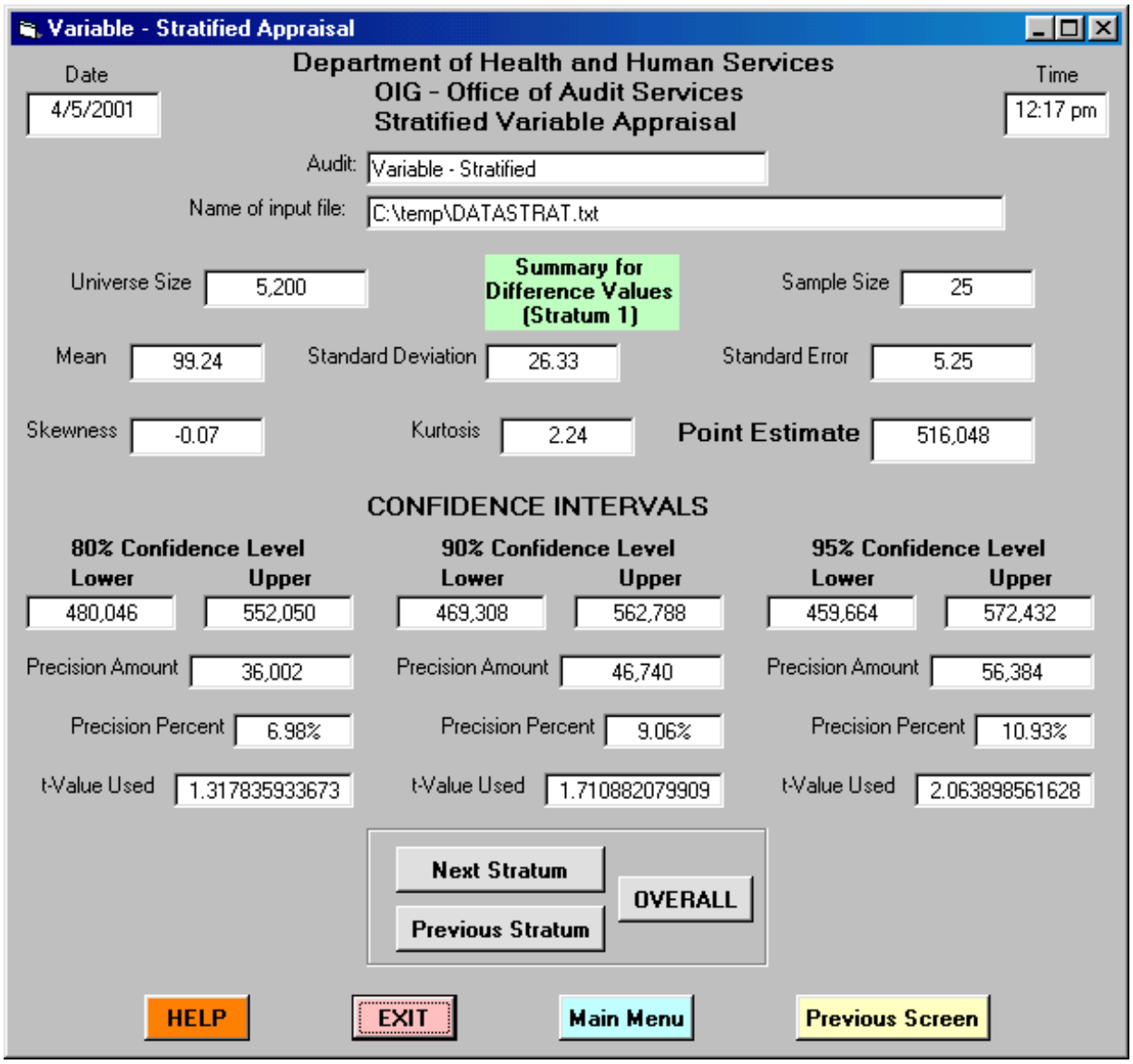
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 4/5/2001          STRATIFIED VARIABLE APPRAISAL              Time: 12:17
                     AUDIT/REVIEW: Variable - Stratified

                    DATA FILE USED: C:\temp\DATASTRAT.txt

          STRATUM     SAMPLE
           NUMBER      SIZE       VALUE OF SAMPLE   NONZERO ITEMS
              1          25              2,481.00              25
              2          25              7,785.00              25

         TOTALS          50             10,266.00              50

            ---------------------- D I F F E R E N C E ----------------------
Stratum 1   MEAN / UNIVERSE                          99.24               5,200
            STANDARD DEVIATION                       26.33
            STANDARD ERROR                            5.25
            SKEWNESS                                  -.07
            KURTOSIS                                  2.24
            POINT ESTIMATE                         516,048

                                        CONFIDENCE LIMITS
                                       80% CONFIDENCE LEVEL
            LOWER LIMIT                              480,046
            UPPER LIMIT                              552,050
            PRECISION AMOUNT                          36,002
            PRECISION PERCENT                           6.98%
            T-VALUE USED                    1.317835933673
```

```
                                  90% CONFIDENCE LEVEL
            LOWER LIMIT                        469,308
            UPPER LIMIT                        562,788
            PRECISION AMOUNT                    46,740
            PRECISION PERCENT                     9.06%
            T-VALUE USED                1.710882079909

                                  95% CONFIDENCE LEVEL
            LOWER LIMIT                        459,664
            UPPER LIMIT                        572,432
            PRECISION AMOUNT                    56,384
            PRECISION PERCENT                     10.93%
            T-VALUE USED                2.063898561628

Stratum 2   MEAN / UNIVERSE                       311.40              3,500
            STANDARD DEVIATION                     39.64
            STANDARD ERROR                          7.90
            SKEWNESS                                -.06
            KURTOSIS                                1.85
            POINT ESTIMATE                     1,089,900

                                   CONFIDENCE LIMITS
                                  80% CONFIDENCE LEVEL
            LOWER LIMIT                      1,053,461
            UPPER LIMIT                      1,126,339
            PRECISION AMOUNT                    36,439
            PRECISION PERCENT                     3.34%
            T-VALUE USED                1.317835933673

                                  90% CONFIDENCE LEVEL
            LOWER LIMIT                      1,042,592
            UPPER LIMIT                      1,137,208
            PRECISION AMOUNT                    47,308
            PRECISION PERCENT                     4.34%
            T-VALUE USED                1.710882079909

                                  95% CONFIDENCE LEVEL
            LOWER LIMIT                      1,032,831
            UPPER LIMIT                      1,146,969
            PRECISION AMOUNT                    57,069
            PRECISION PERCENT                     5.24%
            T-VALUE USED                2.063898561628

OVERALL     POINT ESTIMATE / UNIVERSE        1,605,948              8,700
            STANDARD ERROR                      38,870

                                   CONFIDENCE LIMITS
                                  80% CONFIDENCE LEVEL
            LOWER LIMIT                      1,556,134
            UPPER LIMIT                      1,655,762
            PRECISION AMOUNT                    49,814
            PRECISION PERCENT                     3.10%
```

```
Z-VALUE USED                        1.281551565545

                                 90% CONFIDENCE LEVEL
LOWER LIMIT                                1,542,012
UPPER LIMIT                                1,669,884
PRECISION AMOUNT                              63,936
PRECISION PERCENT                               3.98%
Z-VALUE USED                          1.644853626951

                                 95% CONFIDENCE LEVEL
LOWER LIMIT                                1,529,764
UPPER LIMIT                                1,682,132
PRECISION AMOUNT                              76,184
PRECISION PERCENT                               4.74%
Z-VALUE USED                          1.959963984540
```

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected includes a text file or printer. The screen immediately following is the summary of the difference values for the first stratum in this illustration. If the user created a data file containing two values per sample item (e.g., examined and audited values), the summary for each amount (examined, audited, difference) can be obtained by clicking on **Additional Summary Info** at the bottom of this screen. For this illustration, the **Additional Summary Info** button is not visible in the screen immediately following since only the difference values were contained in the data file. To obtain the results for the second stratum, click on **Next Stratum**. The user can click on **Next Stratum** and **Previous Stratum** to review the results for the individual strata. To obtain the overall results, click on **OVERALL**.


Variable - Stratified Appraisal
Date
4/5/2001
Department of Health and Human Services
OIG - Office of Audit Services
Stratified Variable Appraisal
Time
12:17 pm
Audit: Variable - Stratified
Name of input file: C:\temp\DATASTRAT.txt
Universe Size 5,200
Summary for Difference Values (Stratum 1)
Sample Size 25
Mean 99.24
Standard Deviation 26.33
Standard Error 5.25
Skewness -0.07
Kurtosis 2.24
Point Estimate 516,048
CONFIDENCE INTERVALS
80% Confidence Level
Lower 480,046
Upper 552,050
Precision Amount 36,002
Precision Percent 6.98%
t-Value Used 1.317835933673
90% Confidence Level
Lower 469,308
Upper 562,788
Precision Amount 46,740
Precision Percent 9.06%
t-Value Used 1.710882079909
95% Confidence Level
Lower 459,664
Upper 572,432
Precision Amount 56,384
Precision Percent 10.93%
t-Value Used 2.063898561628
Next Stratum
Previous Stratum
OVERALL
HELP
EXIT
Main Menu
Previous Screen

Sample Size Determination

OVERVIEW

VARIABLE SAMPLE SIZE DETERMINATION
- - **Unrestricted**
- - **Stratified**

ATTRIBUTE SAMPLE SIZE DETERMINATION

The purpose of the sample size determination module is to estimate the necessary sample size for a certain precision at a given confidence level. The program will generate optimum sample sizes for unrestricted and stratified variable samples and attribute samples. A brief example of when to use each module is given below. A detailed explanation of how to use each module is described later in this section.

Variable Sample Size Determination

The Variable Sample Size Determination program allows the user to estimate sample sizes for specified precision percentages and specified confidence levels. In the Variable Unrestricted module, the user will have the option of having the program read a probe sample file to obtain an estimate of the universe mean and standard deviation or input these two estimates directly without reading a probe sample file. The Variable Stratified module will determine sample sizes for situations where the total sample size is both predetermined or unknown.

Attribute Sample Size Determination

The Attribute Sample Size Determination program determines the sample size for an attribute simple random sample. The sample size is determined to provide for a specified degree of precision (using the desired width of the confidence interval) at four levels of confidence (80%, 90%, 95%, and 99%). The resulting sample sizes are the smallest sample sizes capable of meeting the specified precision requirement at the stated confidence level.

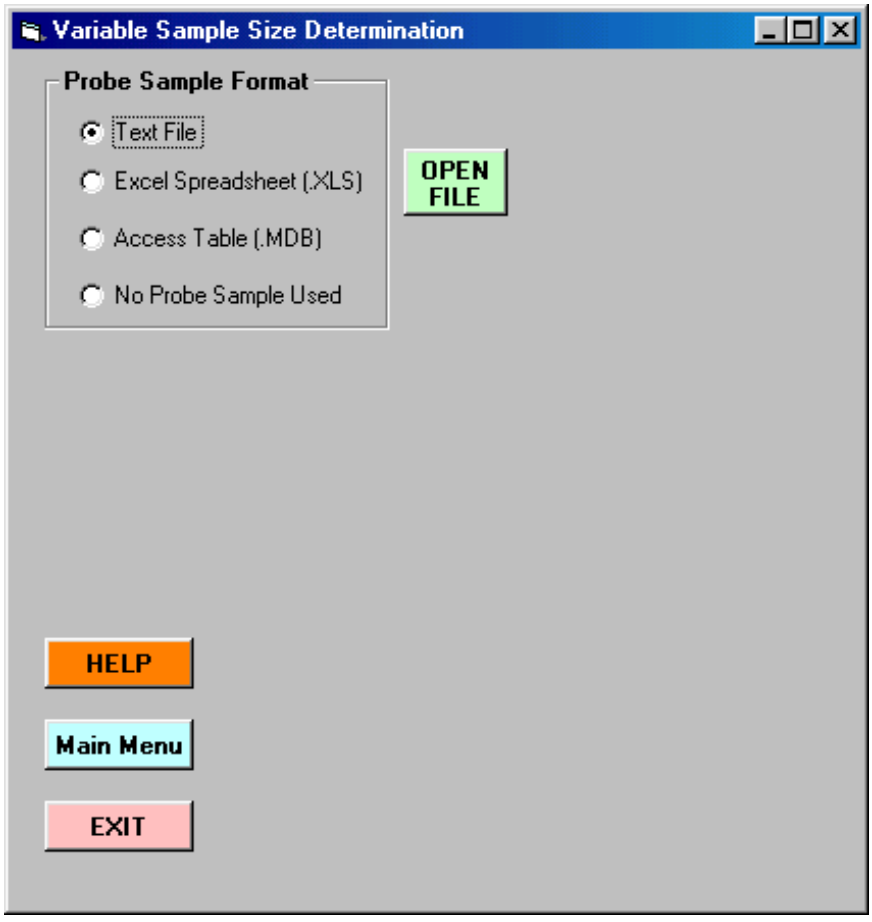
VARIABLE SAMPLE SIZE DETERMINATION UNRESTRICTED

Purpose

This program allows the user to estimate sample sizes for specified precision percentages and specified confidence levels. The user will have the option of having the program read a probe sample file to obtain an estimate of the universe mean and standard deviation or input these two estimates directly without reading a probe sample file.

Input Screen

The input screen for this program is shown below.

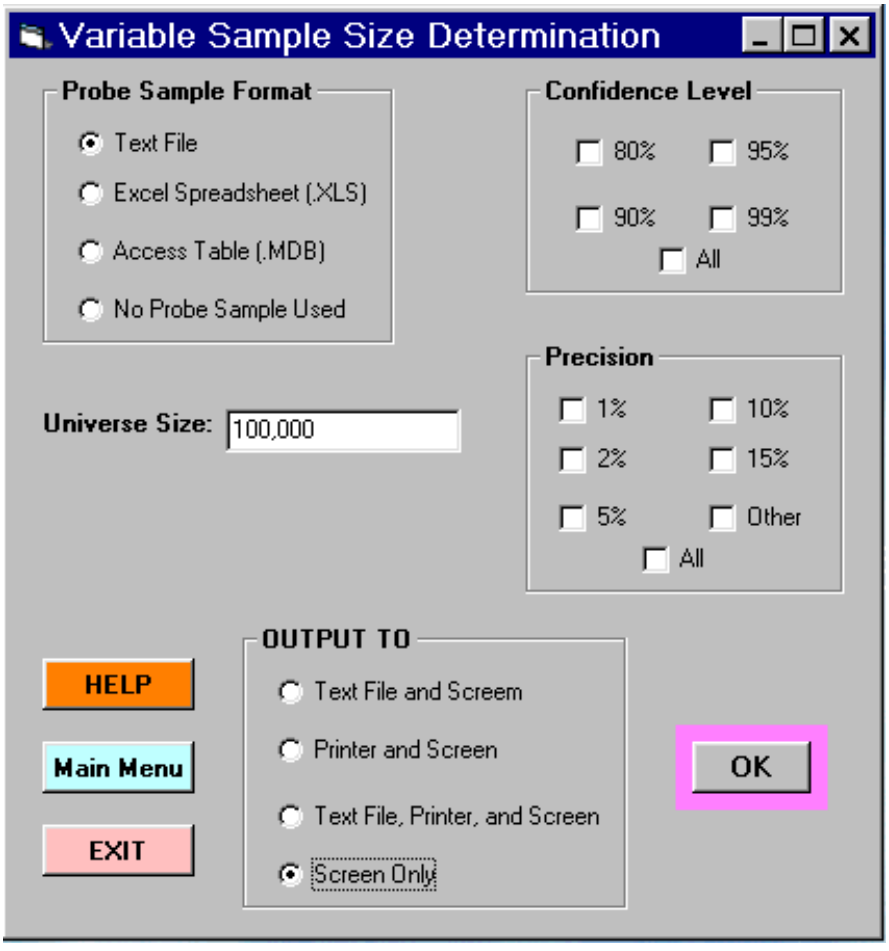


Probe Sample

The user has the option of having the program read a probe sample file to obtain an estimate of the universe mean and standard deviation or inputting these two estimates directly without reading a probe sample file. The probe sample can be contained in a text file, an Excel spreadsheet, or a table within an Access database.

Probe Sample in a Text File

If the probe sample is contained in a text file, click on the **OPEN FILE** button and select the name of this file. By clicking on the **Open** button on the file select form, the user will see the full input screen for this program as shown below.



Full input screen for Sample Size Determination program.

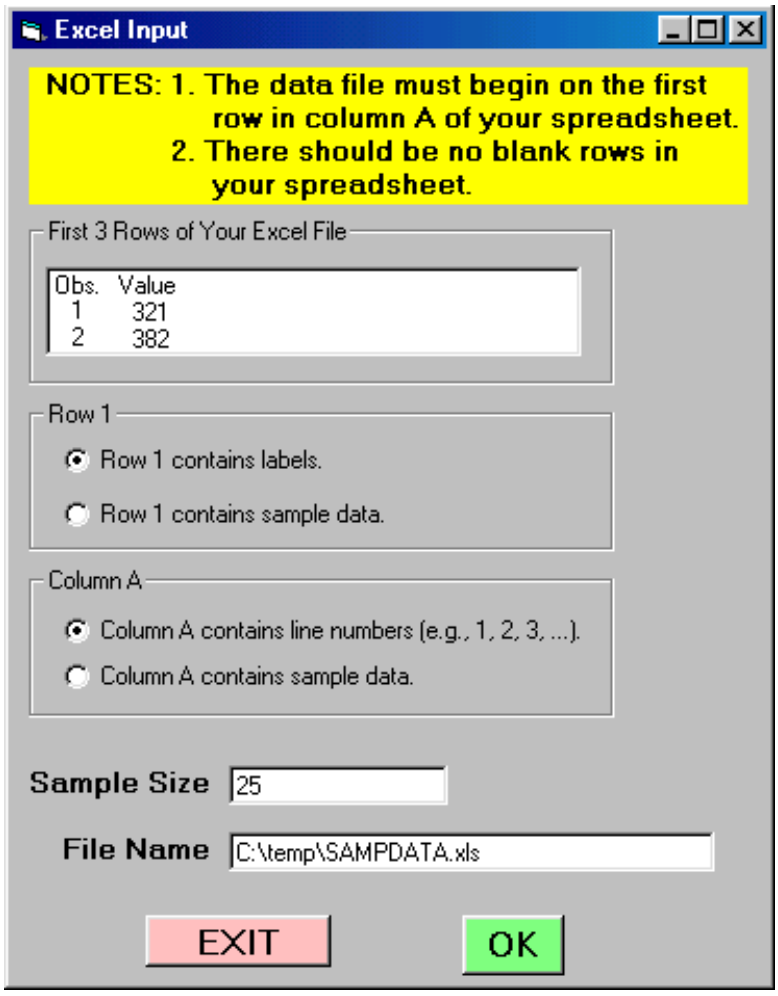
The file used for this illustration is in C:\TEMP\SAMPDATA.txt and contains 25 observations, shown below. This text file should be in a single column with one sample value per line. The mean of this sample is 400 and the standard deviation is 50.

```
321
382
453
459
343
388
313
420
407        The probe text file in C:\TEMP\SAMPDATA.txt.
395
441        (mean = 400, standard deviation = 50)
448
447
333
357
395
477
391
356
368
376
350
461
472
447
```

Probe Sample in an Excel Spreadsheet

If the probe sample is contained in an Excel spreadsheet, click on the **OPEN FILE** button and select the name of this file. By clicking on the **Open** button on the file select form, the user will see the full input screen for this program, shown previously. The Excel file used for this illustration is C:\TEMP\SAMPDATA.XLS and contains the same 25 observations. The mean of this sample is 400 and the standard deviation is 50. This particular file contains labels (variable names) in the first row and contains line numbers (1, 2, 3, ...) in column A. The corresponding options were selected in the Excel Input screen shown next. The line numbers are optional. Had

column A contained the probe sample data, the second option in the Column A frame in the Excel Input screen should have been selected.



	A	B
1	**Obs.**	**Value**
2	1	321
3	2	382
4	3	453
5	4	459
6	5	343
7	6	388
8	7	313
9	8	420
10	9	407
11	10	395
12	11	441
13	12	448
14	13	447
15	14	333
16	15	357
17	16	395
18	17	477
19	18	391
20	19	356
21	20	368
22	21	376
23	22	350
24	23	461
25	24	472
26	25	447

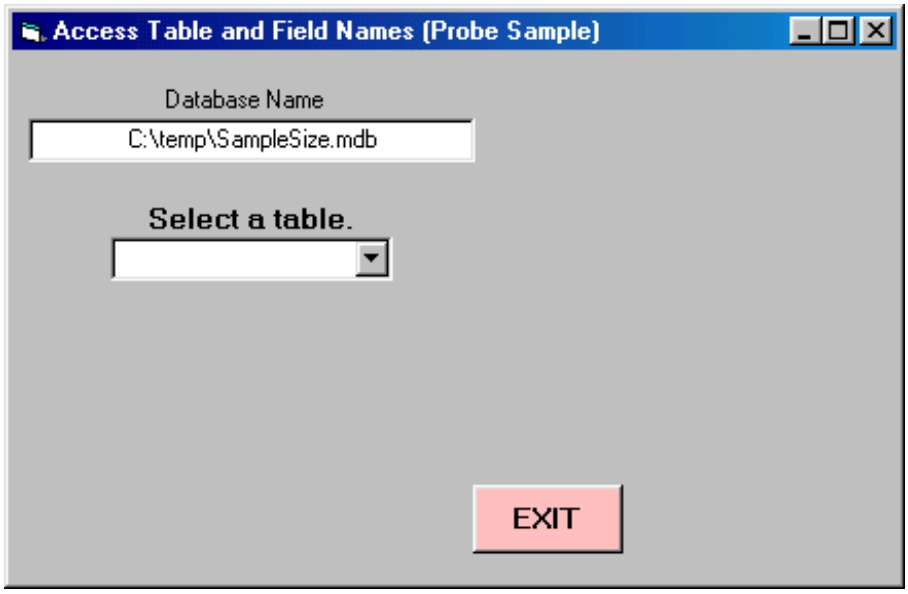
The probe Excel file in C:\TEMP\SAMPDATA.XLS.

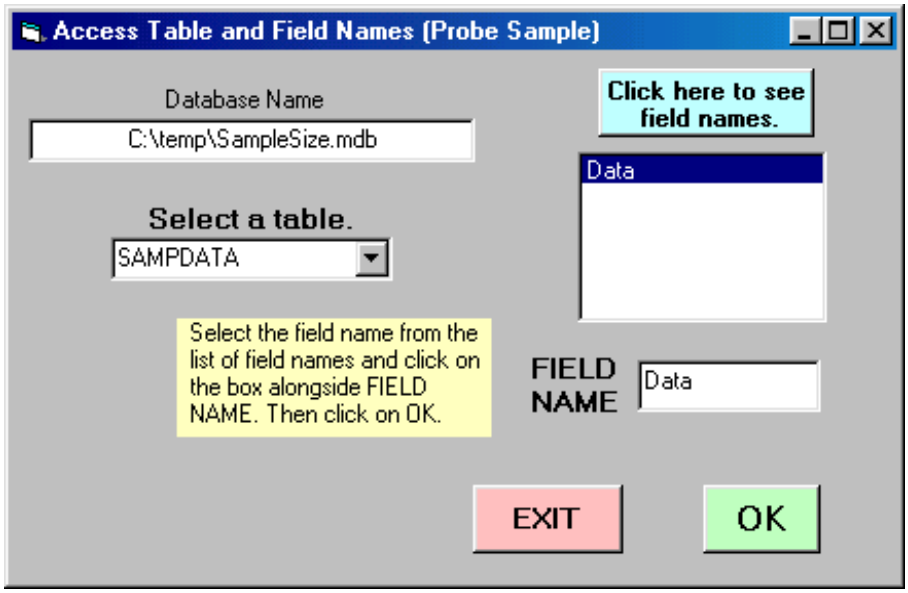
Probe Sample in an Access Database Table

If the probe sample is contained in an Access database table, click on the **OPEN FILE** button and select the name of this database. The Access database used for this illustration is C:\TEMP\SampleSize.mdb and contains a table with 25 observations, the following table shows the first 18 rows. After clicking on the **OPEN FILE** button, the user will be asked to select the name of the table within the selected database using the top form shown immediately after the table. Click on the down arrow under **Select a table**. After selecting a table from the drop-down list (SAMPDATA for this illustration) and clicking on **Click here to see field names**, the form immediately following the probe sample data will appear. To select the field name, click on the field name for the field containing the probe sample (Data in this illustration), then click on the box alongside **FIELD NAME**. Click on **OK** to continue processing.

	Data
▶	321
	382
	453
	459
	343
	388
	313
	420
	407
	395
	441
	448
	447
	333
	357
	395
	477
	391

The first 18 rows of Access table SAMPDATA in Access database C:\TEMP\SampleSize.mdb.





No Probe Sample Used

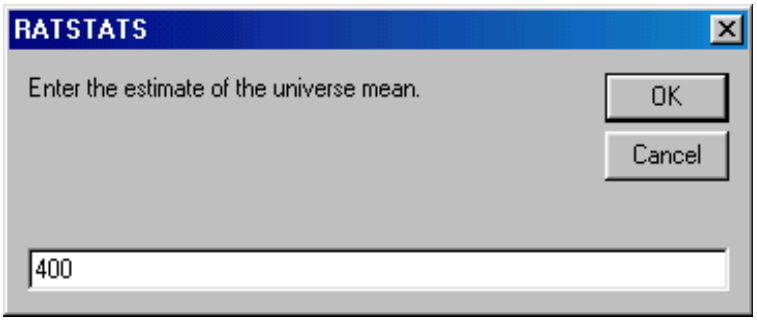
The user has the option of not using a probe sample in this analysis. Using this option, the user will be asked to specify the anticipated mean for the sample. If the mean is not known, the best estimate of the mean may be used. Other sources of data, such as prior reviews, may provide assistance in estimating the mean. Next, the user will be asked to specify the anticipated standard deviation for the sample. This may be the hardest value for the user to approximate. As a guide, approximately two-thirds of the sample values lie between the mean plus or minus one standard deviation. For example, if the user specifies a mean of $400 and a standard deviation of $50, two-thirds of the sample values should lie between $350 and $450. Also, nearly all the sample values should lie within plus or minus three standard deviations of the sample mean (between $250 and $550 for this example).

When the standard deviation is not known, the user has several alternatives for approximating it. Statistical Auditing by Donald Roberts includes several methods for approximating the sample standard deviation.

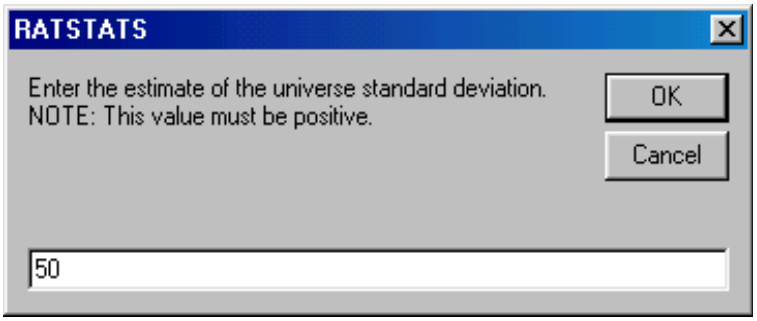
When this option is selected the user will see the following message.

By clicking on **OK**, the user will return to the full input screen for this program, shown previously.

Later in the program, the user will be asked to enter the estimated mean in the box shown below.



Also, the user will be asked to enter the estimated standard deviation in a similar input box.



Working with the Full Input Screen

Confidence Level

The user may select any combination of the following confidence levels: 80%, 90%, 95%, and 99%. Selecting all four confidence levels can be done by clicking on the "All" option. If none of the options are selected, the program will use the default "All" option.

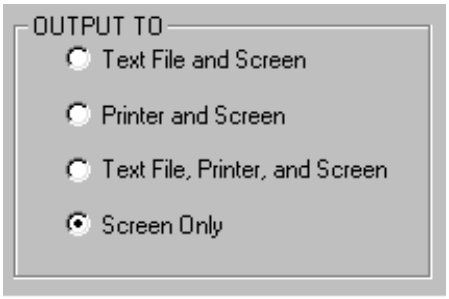
Precision

Sample precision is measured as a percentage. The available precision percentages are 1%, 2%, 5%, 10%, 20%, and "Other." By selecting "All," the user will obtain the sample sizes for the first five precision percentages. When selecting "Other," the user will be prompted to enter the desired precision percentage. Enter this value as an integer, (e.g., 25 for 25%, 35 for 35%). After entering this value, on the screen the user will see the word "Other" change to the specified value. If none of the options are selected, the program will use the default "All" option.

A confidence interval for the universe mean is obtained by adding and subtracting an amount (say, E) to/from the sample mean. The value of E is determined using the universe size, sample size, standard deviation and the selected confidence level. The sample precision percentage is calculated as E divided by the point estimate of the universe mean times 100.

Program Output

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by clicking the corresponding button.



NOTE: The program always concludes with a summary on the screen.

Explanation of Output

The output for each cell in the output table will consist of (1) the necessary sample size or (2) the text "- - -". The necessary sample size is the number of sample items necessary to obtain the specified sample precision at the specified confidence level. For example, in this illustration, a sample size of 106 is necessary to obtain a point estimate having a precision percentage of plus or minus 2% using a 90% confidence level. If the calculated sample size is zero, a text value of "- - -" will appear in this cell. This occurred in the lower left cell for the sample illustration.

The output also contains the estimated mean and standard deviation, along with the specified universe size.

Output to a Text File or Printer

Using the probe sample in C:\TEMP\SAMPDATA.TXT, the sample sizes were saved in C:\TEMP\OUTSIZES.TXT, shown below. The printer output is identical. If any of the sample sizes are under 30, the note shown following the calculated sample sizes is the final portion of the program output.

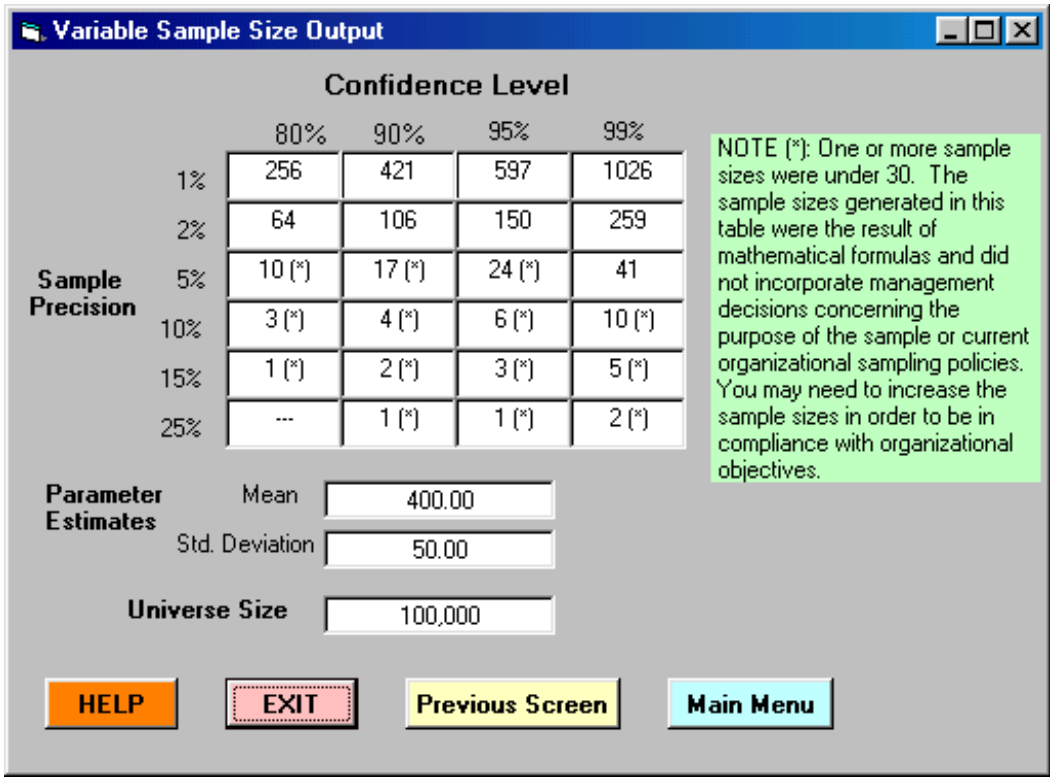
```
                DEPARTMENT OF HEALTH & HUMAN SERVICES
                    OIG - OFFICE OF AUDIT SERVICES
Date: 5/11/2000      Sample Size Determination          Time: 21:52


                                  Confidence Level
                      80%          90%            95%            99%
            1%        256          421            597            1026
            2%        64           106            150            259
Precision   5%        10 (*)       17 (*)         24 (*)         41
Level      10%        3 (*)        4 (*)          6 (*)          10 (*)
           15%        1 (*)        2 (*)          3 (*)          5 (*)
           25%        ---          1 (*)          1 (*)          2 (*)


Estimated Mean: 400.00

Estimated Std. Deviation: 50.00

Universe Size: 100,000
```

Output - continued

```
NOTE (*): One or more sample sizes were under 30.  The generated
          sample sizes were the result of mathematical formulas
          and did not incorporate management decisions concerning
          the purpose of the sample or current organizational
          sampling policies.  You may need to increase the sample
          sizes in order to be in compliance with organizational
          objectives.
```

Output to Screen

The program always concludes with a screen summary. The screen output for this illustration is shown below. A universe size of 100,000 is used and the option "All" was selected for the confidence levels and precision percentages. The "Other" precision percentage was specified as 25%. The note on the right side of this screen will appear whenever one or more of the sample sizes are under 30.



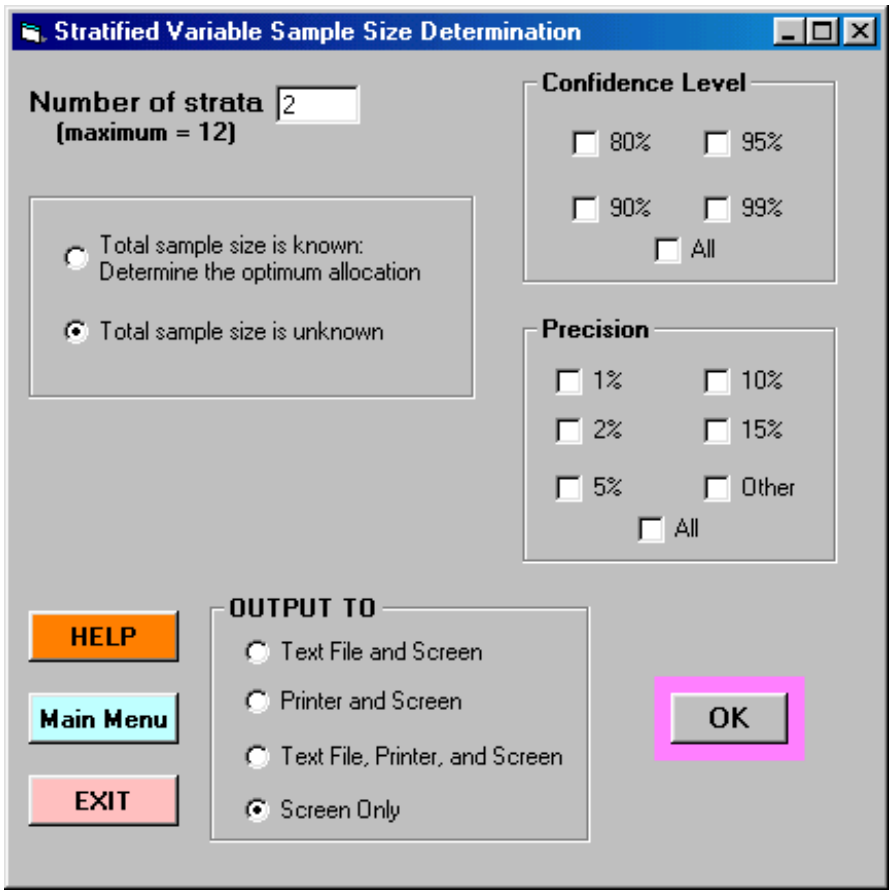
VARIABLE SAMPLE SIZE DETERMINATION STRATIFIED

Purpose

This program allows the user to estimate sample sizes for specified precision percentages and specified confidence levels. This program will generate sample sizes for stratified samples. The program allows for sample sizes for up to 12 strata. The total sample size may be determined by the program or specified by the user.

Input Screen

The input screen for this program is shown below.



Sample Size is Known / Sample Size is Unknown

If the user plans on performing a stratified sample and the overall sample size has been predetermined, click on "Sample size is known: Determine the optimal allocation." Enter the total sample size in the box shown below.

If the overall sample size has not been predetermined, click on "Total sample size is unknown."

Working with the Full Input Screen

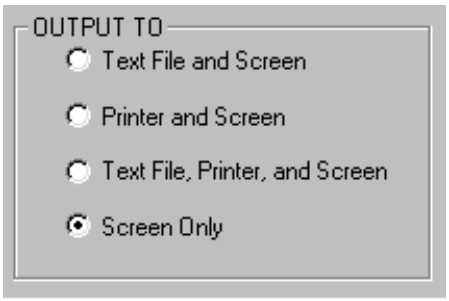
Confidence Level

The user may select any combination of the following confidence levels: 80%, 90%, 95%, and 99%. Selecting all four confidence levels can be done by clicking on the "All" option. If none of the options are selected, the program will use the default "All" option.

Precision

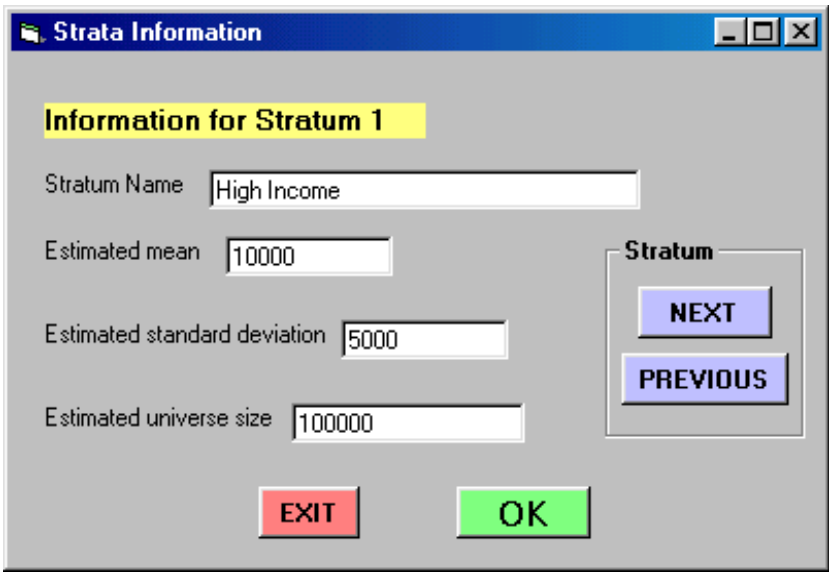
Sample precision is measured as a percentage. The available precision percentages are 1%, 2%, 5%, 10%, 20%, and "Other." By selecting "All," the user will obtain the sample sizes for the first five precision percentages. When selecting "Other," the user will be prompted to enter the desired precision percentage. Enter this value as an integer, (e.g., 25 for 25%, 35 for 35%). After entering this value, on the screen the user will see the word "Other" change to the specified value. If none of the options are selected, the program will use the default "All" option.

Program Output



The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The program always concludes with a summary on the screen.

After clicking on **OK**, the user will need to specify information for each of the strata using the following form. To move to the next stratum, click on **NEXT**. The input values for each stratum may be viewed by clicking on **NEXT** and **PREVIOUS** to move from one stratum to another. When all input values have been entered, click on **OK**.



Explanation of Output

The output for each cell in the output table will consist of (1) the necessary sample size or (2) the text "- - -". The necessary sample size is the number of sample items necessary to obtain the specified sample precision at the specified confidence level. For example, in this illustration, in the first stratum, a sample size of 110 is necessary to obtain a point estimate having a precision percentage of plus or minus 5% using a 90% confidence level. If the calculated sample size is zero, a text value of "- - -" will appear in this cell.

The output also contains the user-specified (1) estimated mean, (2) estimated standard deviation, and (3) estimated universe size. The computed ratio of the total sample size allocated to this stratum is also contained in the output.

Program Output - Total Sample Size is Unknown

Output to a Text File or Printer

The sample sizes for this illustration were saved in C:\TEMP\OUTSTRSIZES.TXT, shown below. The printer output is identical. The option "All" was selected for the confidence levels and precision percentages and the "Other" precision percentage was specified as 25%.

NOTES: (1) The program calculates the estimated mean and standard deviation of the entire universe ($5,8333.33 and $4,579.54, respectively, in this illustration). (2) Whenever one or more of the sample sizes are under 30, the program output will conclude with the note immediately following the calculated total sample sizes.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 12/19/2000            Sample Size Determination              Time: 12:02


                THE ESTIMATES ARE BASED ON THE FOLLOWING ENTRIES:

NBR   DESCRIPTION       -- MEAN --   -- STD.DEV. --   -- UNIVERSE --   -- RATIO --
  1   High Income        10,000.00         5,000.00          100,000        20.00%
  2   Low Income          5,000.00         4,000.00          500,000        80.00%

- TOTALS -                5,833.33         4,579.54          600,000
= = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
Sample Sizes for Stratum 1: High Income

                                Confidence Level
                  80%           90%             95%             99%
            1%    1653          2699            3795            6406
            2%    418           687             972             1669
Precision   5%    67            111             157             271
Level      10%    17 (*)        28 (*)          40              68
           15%    8 (*)         13 (*)          18 (*)          31
           25%    3 (*)         5 (*)           7 (*)           11 (*)

Sample Sizes for Stratum 2: Low Income

                                Confidence Level
                  80%           90%             95%             99%
            1%    6611          10793           15180           25624
            2%    1671          2745            3888            6676
Precision   5%    268           442             627             1081
Level      10%    68            111             157             271
           15%    30            50              70              121
           25%    11 (*)        18 (*)          26 (*)          44
```

Total Sample Sizes

		Confidence Level			
		80%	90%	95%	99%
Precision Level	1%	8264	13492	18975	32030
	2%	2089	3432	4860	8345
	5%	335	553	784	1352
	10%	85	139	197	339
	15%	38	63	88	152
	25%	14 (*)	23 (*)	33	55

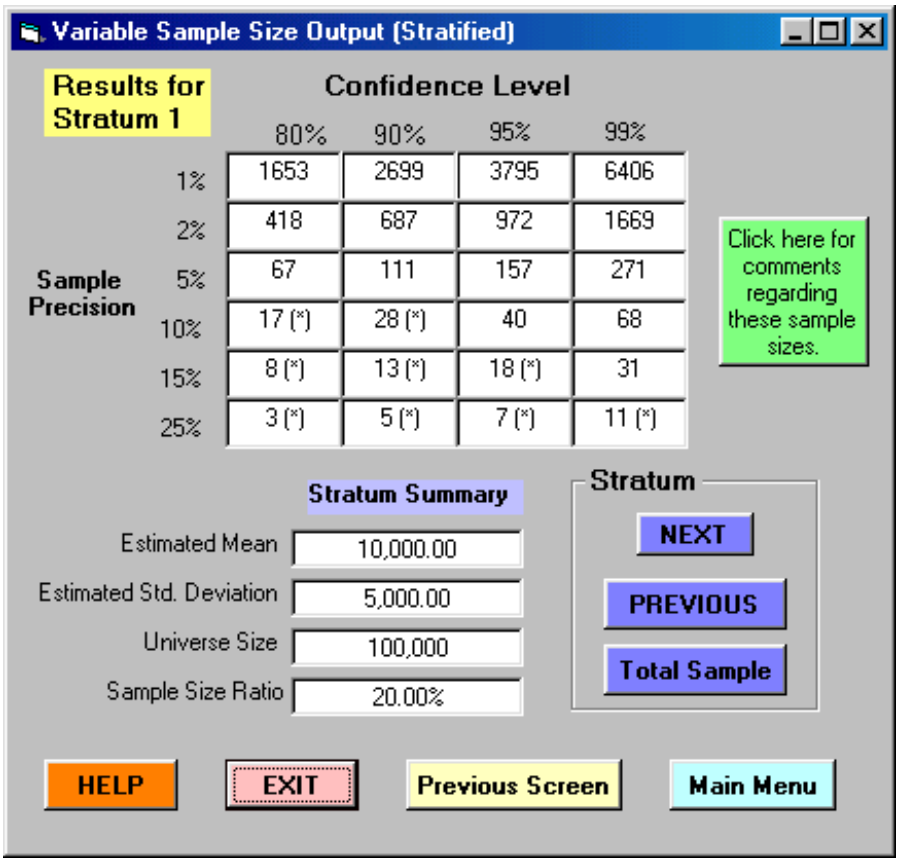
NOTE (*): One or more sample sizes were under 30. The generated sample sizes were the result of mathematical formulas and did not incorporate management decisions concerning the purpose of the sample or current organizational sampling policies. You may need to increase the sample sizes in order to be in compliance with organizational objectives.

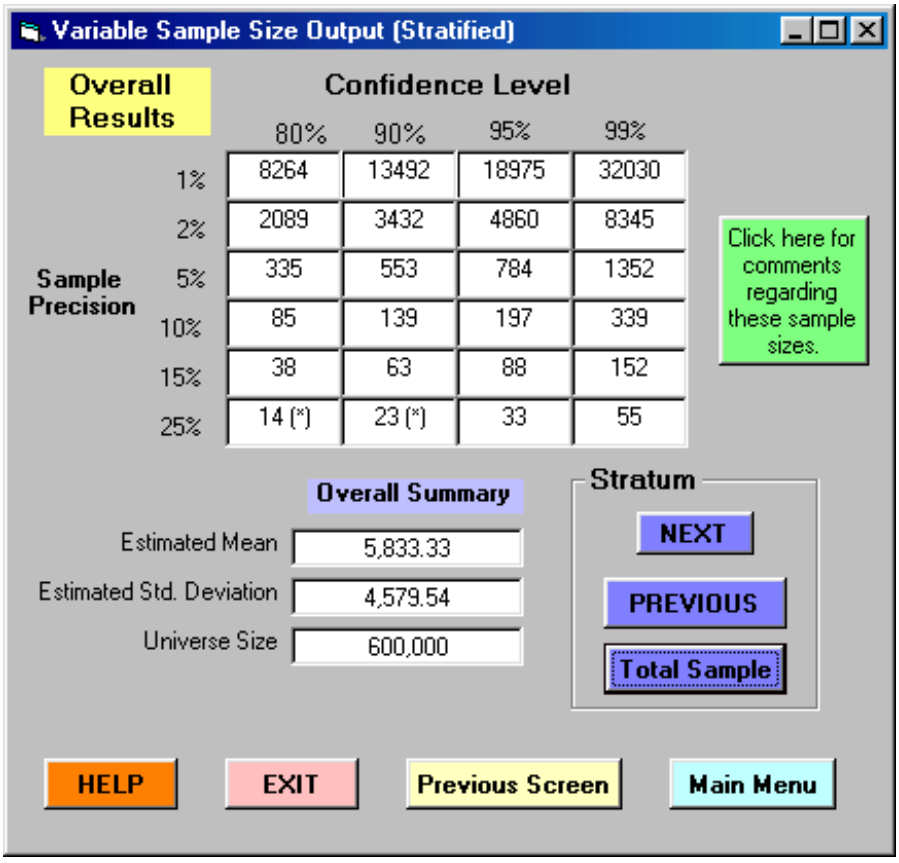
If any of the calculated samples sizes exceeds the corresponding universe size, the program will conclude with the following reminder.

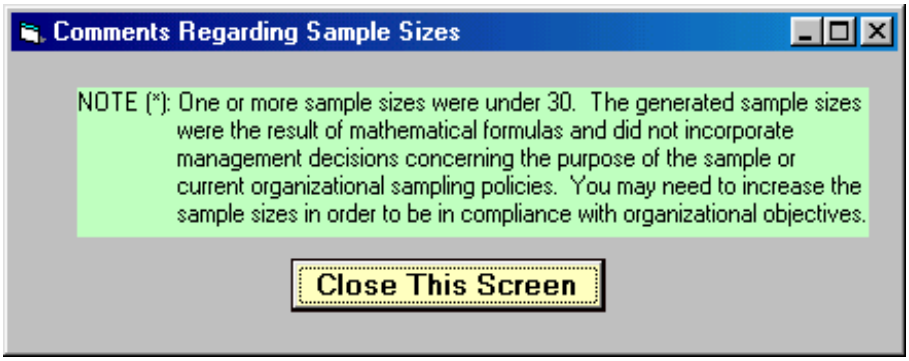
NOTE (#): The formulas calculated a sample size greater than the universe size. The program reduced the calculated sample size to the universe size. The additional sampling units were then distributed among the remaining strata based on optimal allocation formulas.

Output to Screen

The program always concludes with a screen summary. The screen output for the first stratum in this illustration is shown next. The option "All" was selected for the confidence levels and precision percentages. The "Other" precision percentage was specified as 25%. To view the output for the second stratum, click on **NEXT** and to see the overall results (total sample sizes), click on **Total Sample**. The window that appears when clicking on "Click here for comments regarding these sample sizes" immediately follows the two output screens. If any of the calculated samples sizes exceeds the corresponding universe size, the comments window will also contain NOTE (#) from the previous page.







NOTE: The program calculates the estimated mean and standard deviation of the entire universe ($5,8333.33 and $4,579.54, respectively, in this illustration).

Program Output - Total Sample Size is Known

Output to a Text File or Printer

The sample sizes for this illustration were saved in C:\TEMP\OUTSTRKNOWN.TXT, shown next. The printer output is identical. The total sample size was specified as 500. The option "All" was selected for the confidence levels.

NOTE: The program calculates the estimated mean and standard deviation of the entire universe ($5,8333.33 and $4,579.54, respectively, in this illustration). The calculated sample sizes are 100 (High Income stratum) and 400 (Low Income stratum).

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 12/19/2000            Sample Size Determination              Time: 13:07

         THE ESTIMATES ARE BASED ON THE FOLLOWING ENTRIES:

NBR   DESCRIPTION       -- MEAN --  -- STD.DEV. --  -- UNIVERSE --
  1   High Income        10,000.00        5,000.00         100,000
  2   Low Income          5,000.00        4,000.00         500,000

- TOTALS -                5,833.33        4,579.54         600,000
= = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =

Precision Values:

Confidence Level    80%         90%           95%           99%
                  4.09%       5.25%         6.26%         8.22%

The following sample sizes are based on a total sample size of 500.

Stratum 1: High Income
  Sample
    Size      Ratio
     100     20.00%

Stratum 2: Low Income
  Sample
    Size      Ratio
     400     80.00%
```

If any of the sample sizes are under 30 the text file/printer output will contain the following reminder.

```
NOTE (*):   One or more sample sizes were under 30.  The generated
            sample sizes were the result of mathematical formulas
            and did not incorporate management decisions
            concerning the purpose of the sample or current
            organizational sampling policies.  You may need to
            increase the sample sizes in order to be in compliance
            with organizational objectives.
```

If any of the calculated samples sizes exceeds the corresponding universe size, the program output will contain the following reminder.

```
NOTE (#):   The formulas calculated a sample size greater than the
            universe size.  The program reduced the calculated
            sample size to the universe size.  The additional
            sampling units were then distributed among the
            remaining strata based on optimal allocation formulas.
```

If, due to rounding, the total sample size calculated does not equal the total sample size requested, the program will contain a reminder similar to the following:

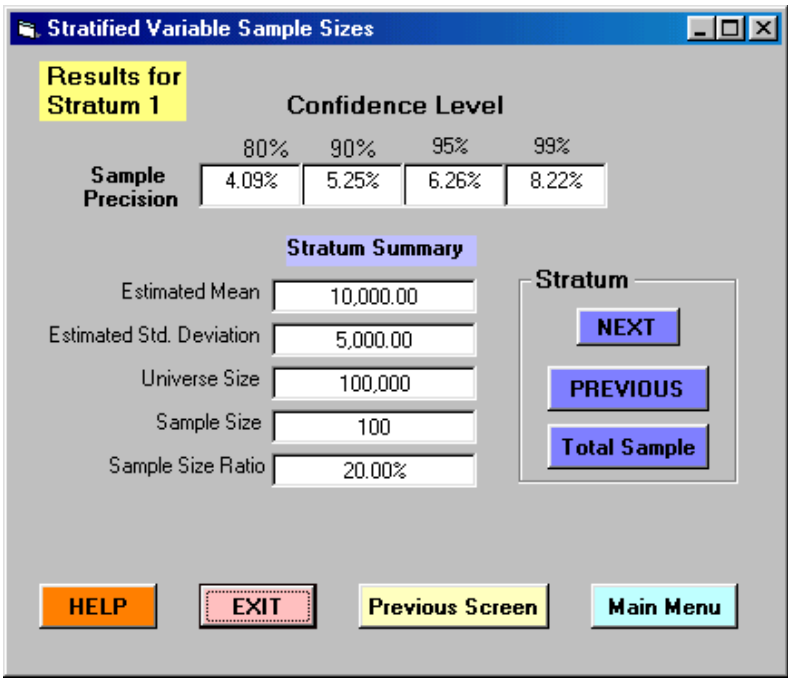
```
NOTE (!):   Due to rounding, the total sample size calculated
            (499) does not equal the sample size requested (500).
```

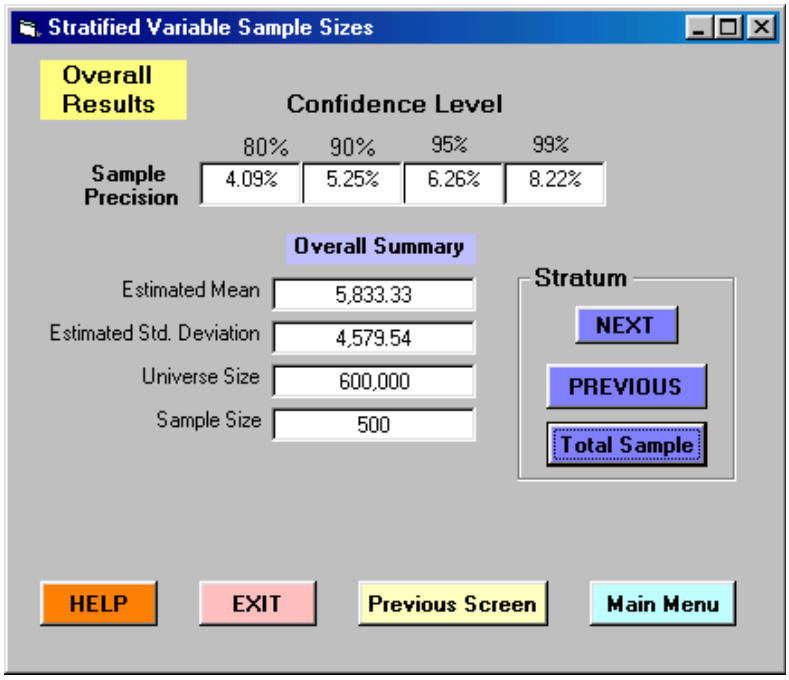
Output to Screen

The program always concludes with a screen summary. The screen output for the first stratum in this illustration is shown next. The total sample size was specified as 500. The option "All" was selected for the confidence levels. To view the output for the second stratum, click on **NEXT** and to see the overall results (total sample sizes), click on **Total Sample**.

NOTE: The program calculates the estimated mean and standard deviation of the entire universe ($5,8333.33 and $4,579.54, respectively, in this illustration). The calculated sample sizes are 100 (High Income stratum) and 400 (Low Income stratum).

If any of the conditions described in NOTES (*), (#), or (!) above exist, the screen output will contain the following button. By clicking on this button, the user will see the corresponding notes, depending on the existing conditions.


Stratified Variable Sample Sizes
Results for Stratum 1
Confidence Level
80%
90%
95%
99%
Sample Precision
4.09%
5.25%
6.26%
8.22%
Stratum Summary
Estimated Mean
10,000.00
Estimated Std. Deviation
5,000.00
Universe Size
100,000
Sample Size
100
Sample Size Ratio
20.00%
Stratum
NEXT
PREVIOUS
Total Sample
HELP
EXIT
Previous Screen
Main Menu


Stratified Variable Sample Sizes
Overall Results
Confidence Level
80%
90%
95%
99%
Sample Precision
4.09%
5.25%
6.26%
8.22%
Overall Summary
Estimated Mean
5,833.33
Estimated Std. Deviation
4,579.54
Universe Size
600,000
Sample Size
500
Stratum
NEXT
PREVIOUS
Total Sample
HELP
EXIT
Previous Screen
Main Menu

Attribute Sample Size Determination

Purpose

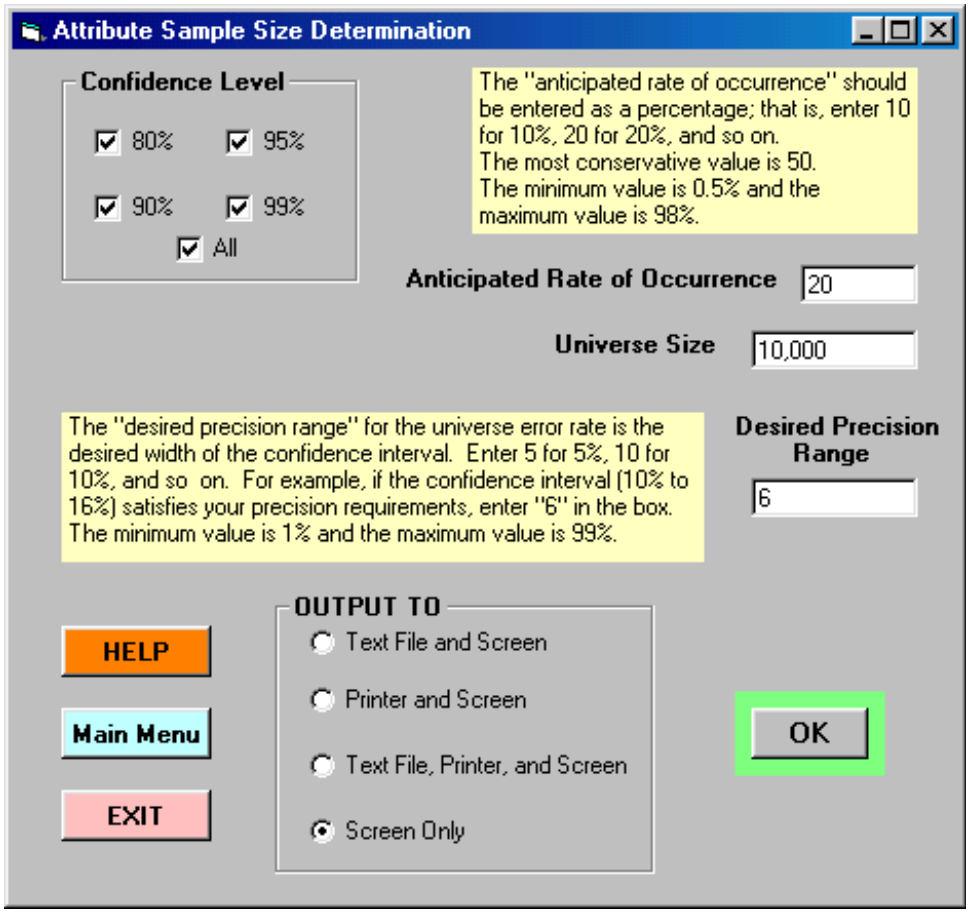
This program determines the sample size for an attribute simple random sample. The sample size is determined to provide for a specified degree of precision (using the desired width of the confidence interval) at four levels of confidence (80%, 90%, 95%, and 99%). The resulting sample sizes are the smallest sample sizes capable of meeting the specified precision requirement at the stated confidence level.

Confidence intervals for attribute sampling are exact and are based on the hypergeometric distribution. As a result, such confidence intervals are usually not symmetric about the point estimate. For example, the point estimate may be 3% and the corresponding 90% confidence interval is from 2% to 6%. For this illustration, the *width of the confidence interval* is 4% and the *confidence level* is 90%. Consequently, attribute confidence intervals differ from the usual interval obtained by deriving the point estimate plus or minus the estimated precision, where the estimated precision is half the width of the resulting confidence interval. Because of this, the "desired precision" for the attribute sampling procedure must be specified as the desired width (rather than the half width) of the confidence interval.

The input requirements also include (1) the size of the universe and (2) the anticipated rate of occurrence in the universe. This rate of occurrence is generally estimated from past experience, either from similar systems or a past review of this universe. If no information concerning the rate of occurrence is available, the most conservative procedure is to specify 50% for this value. If the actual rate of occurrence differs from the user-specified rate of occurrence, this in no way affects the sample's validity, but the resulting precision (confidence interval width) will likely differ from the specified "desired precision."

Input Screen

The input screen for this program is shown next.



Input Values

Confidence Level

The user may select any combination of the following confidence levels: 80%, 90%, 95%, and 99%. Selecting all four confidence levels can be done by clicking on the "All" option.

Anticipated Rate of Occurrence

This value is the expected rate of occurrence for the universe. It should be expressed as a percentage (e.g., enter "15" not ".15"). If no information is available for the anticipated rate of occurrence, the most conservative procedure is to specify a rate of occurrence of 50%.

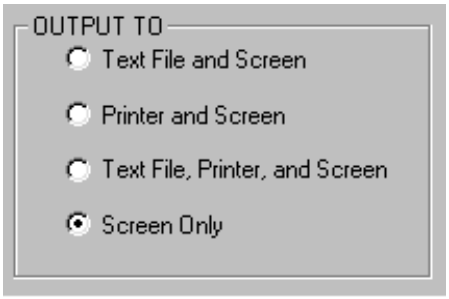
Universe Size

The universe size is the total number of items from which the sampled items were selected. This number should be entered without commas (e.g., 50000, not 50,000). The program will insert commas upon leaving this box.

Desired Precision Range

This value is the desired width of the resulting confidence interval, expressed as a percentage. It is equal to the upper confidence limit minus the lower limit. For example, if the confidence interval (10% to 16%) satisfies the user's precision requirements, the response to this query would be "6".

Program Output



The output options are: (1) a text file and screen, (2) a printer and screen, (3) a text file, printer and screen, or (4) screen only. The program always finishes with the screen output.

Explanation of Output

The output for each cell in the output table will consist of (1) the necessary sample size or (2) the text "- - - ". The necessary sample size is the number of sample items necessary to obtain the specified sample precision at each confidence level. For example, in this illustration a sample size of 488 is necessary to obtain a confidence interval having a width of 6% using a 90% confidence level. If the calculated sample size is zero, a text value of "- - -" will appear in this cell.

The output also contains the user specified anticipated rate of occurrence, desired precision range, and universe size.

Output to a Text File or Printer

The sample sizes for this illustration were saved in C:\TEMP\OUTSIZES.TXT, shown below. The printer output is identical. The input values are those shown on the initial input screen where the option "All" was selected for the confidence levels.

```
                DEPARTMENT OF HEALTH & HUMAN SERVICES
                   OIG - OFFICE OF AUDIT SERVICES
Date: 12/19/2000   Sample Size Determination          Time: 8:46


                              Confidence Level
                  80%          90%           95%           99%
Sample Size       314          488           666         1,079


Anticipated Rate of Occurrence: 20%

Desired Precision Range: 6%

Universe Size: 10,000
```

If any of the samples sizes are under 30, the following note will be the final part of the text file and/or printer output.

```
NOTE (*): One or more sample sizes were under 30.  The generated
          sample sizes were the result of mathematical formulas
          and did not incorporate management decisions concerning
          the purpose of the sample or current organizational
          sampling policies.  You may need to increase the sample
          sizes in order to be in compliance with organizational
          objectives.
```

Output to Screen

The program always concludes with a screen summary. The screen output for this illustration is shown below. If one or more sample sizes are under 30, these sample sizes are flagged using "(*)" (e.g., 22 (*)) and the following note will appear in the output screen.

NOTE (*): One or more sample sizes were under 30. The generated sample sizes were the result of mathematical formulas and did not incorporate management decisions concerning the purpose of the sample or current organizational sampling policies. You may need to increase the sample sizes in order to be in compliance with organizational objectives.