RAT-STATS 2007 User Guide

Version 2

Department of Health & Human Services
Office of Inspector General
Office of Audit Services

TABLE OF CONTENTS

Page

GETTING STARTED

Introduction
Installation i
Frequently Asked Questions ii

Windows Version of RAT-STATS
Opening Screen 1-1
Random Numbers 1-2
Attribute Appraisals 1-3
Variable Appraisals 1-4
Sample Size Determination 1-5

RANDOM NUMBERS

Overview
Single Stage Numbers 2-2
Sets of Two Numbers 2-2
Sets of Three Numbers 2-2
Sets of Four Numbers 2-2
Frames - Single Stage 2-2
Frames - Sets of Two 2-3
RHC Sample Selection 2-3

Single Stage Numbers
Purpose 2-4
Input Screen 2-4
Output Options 2-6
Program Output 2-6
Output to a Text File 2-6
Output to an Access Database 2-8
Output to an Excel Spreadsheet 2-9
Output to a Flat File 2-10
Output to a Printer 2-12
Output to Screen 2-13

Sets of Two Numbers

Purpose 2-14
Input Screen 2-14
Output Options 2-16
Program Output 2-16
Output to a Text File 2-17
Output to an Access Database 2-18
Output to an Excel Spreadsheet 2-19
Output to a Flat File 2-21
Output to a Printer 2-22
Output to Screen 2-24

Sets of Three Numbers

Purpose 2-25
Input Screen 2-25
Output Options 2-27
Program Output 2-27
Output to a Text File 2-28
Output to an Access Database 2-29
Output to an Excel Spreadsheet 2-30
Output to a Flat File 2-31
Output to a Printer 2-32
Output to Screen 2-34

Sets of Four Numbers

Purpose 2-35
Input Screen 2-35
Output Options 2-37
Program Output 2-37
Output to a Text File 2-38
Output to an Access Database 2-39
Output to an Excel Spreadsheet 2-40
Output to a Flat File 2-41
Output to a Printer 2-42
Output to Screen 2-44

Frames - Single Stage

Purpose 2-45
Input Screen 2-45
Output Options 2-47
Program Output 2-47
Output to a Text File 2-48
Output to an Access Database 2-49

Output to an Excel Spreadsheet 2-50
Output to a Flat File 2-51
Output to a Printer 2-52
Output to Screen 2-54

Frames - Sets of Two Numbers
Purpose 2-55
Input Screen 2-55
Output Options 2-57
Program Output 2-58
Output to a Text File 2-58
Output to an Access Database 2-59
Output to an Excel Spreadsheet 2-61
Output to a Flat File 2-62
Output to a Printer 2-63
Output to Screen 2-65

RHC Sample Selection
Purpose 2-66
Input Screen for Selecting Primary/Secondary Units 2-67
Selecting the Number of Primary/Secondary Units to be Sampled 2-67
Seed Values 2-68
Input File Information 2-68
Format of Input File 2-69
Input from a Text File 2-70
Input from an Access Database 2-71
Input from an Excel Spreadsheet 2-73
Output Information 2-73
Format for Complete Output 2-74
Format for Summary Output 2-74
Text Summary File 2-75
Access Summary File 2-75
Excel Summary File 2-75
Program Output 2-76
Complete Output to Text File or Printer 2-77
Summary Output to Text File 2-79
Summary Output to Access Database 2-80
Summary Output to Excel Spreadsheet 2-80
Summary Output to Screen 2-81

ATTRIBUTE APPRAISALS

Overview
Unrestricted 3-2
Stratified 3-2
Two-Stage Unrestricted 3-2
Three-Stage Unrestricted 3-2
RHC Two-Stage 3-2
RHC Three-Stage 3-3
Stratified Cluster 3-3
Stratified Multistage 3-3

Unrestricted
Purpose 3-4
Input Screen 3-4
Output Options 3-5
Program Output 3-6
Output to a Text File or Printer 3-7
Output to Screen 3-10

Stratified
Purpose 3-12
Input Screen 3-12
Entering the Sample Results from the Screen 3-14
Output Options 3-16
Program Output 3-16
Output to a Text File or Printer 3-18
Output to Screen 3-19

Two-Stage Unrestricted
Purpose 3-20
Input Screen 3-20
Entering the Sample Results from the Screen 3-24
Output Options 3-25
Program Output 3-25
Output to a Text File or Printer 3-27
Output to Screen 3-28

Three-Stage Unrestricted
Purpose 3-29
Input Screen 3-29
Output Options 3-34
Program Output 3-34

Output to a Text File or Printer 3-36
Output to Screen 3-38

RHC Two Stage
Purpose 3-39
Input Screen 3-39
Output Options 3-43
Program Output 3-43
Output to a Text File or Printer 3-47
Output to Screen 3-49

RHC Three Stage
Purpose 3-50
Input Screen 3-50
Output Options 3-56
Program Output 3-57
Output to a Text File or Printer 3-60
Output to Screen 3-63

Stratified Cluster
Purpose 3-64
Input Screen 3-64
Output Options 3-69
Program Output 3-69
Output to a Text File or Printer 3-71
Output to Screen 3-73

Stratified Multistage
Purpose 3-74
Input Screen 3-74
Entering the Sample Results from the Screen 3-77
Output Options 3-77
Program Output 3-78
Output to a Text File or Printer 3-79
Output to Screen 3-80

VARIABLE APPRAISALS

Overview
Unrestricted 4-2
Stratified 4-2
Two-Stage Unrestricted 4-2

Three-Stage Unrestricted 4-2
RHC Two-Stage 4-2
RHC Three-Stage 4-3
Stratified Cluster 4-3
Stratified Multistage 4-3
Poststratification 4-3
Unknown Universe Size 4-3

Unrestricted
Purpose 4-5
Input Screen 4-5
Format of Input File 4-7
Input from a Text File 4-8
Input from an Access Database 4-9
Input from an Excel Spreadsheet 4-12
Output Options 4-13
Program Output 4-14
Output to a Text File or Printer 4-15
Output to Screen 4-18

Stratified
Purpose 4-19
Input Screen 4-19
Format of Input File 4-21
Input from a Text File 4-22
Input from an Access Database 4-25
Input from an Excel Spreadsheet 4-28
Complete or Summary Output 4-29
Output Options 4-30
Program Output 4-31
Output to a Text File or Printer 4-32
Output to Screen 4-35

Two-Stage Unrestricted
Purpose 4-37
Input Screen 4-37
Format of Input File 4-39
Input from Text Files 4-40
Input from an Access Database 4-43
Input from Excel Spreadsheets 4-46
Specifying the Number of Secondary Units 4-48
Output Options 4-49

Program Output 4-50
Output to a Text File or Printer 4-51
Output to Screen 4-52

Three-Stage Unrestricted
Purpose 4-55
Input Screen 4-55
Format of Input File 4-57
Format of Primary Unit Information 4-58
Format of Secondary Unit Information 4-58
Format of Sample Data Information 4-59
Input from Text Files 4-60
Input from an Access Database 4-63
Input from Excel Spreadsheets 4-66
Complete or Summary Output 4-69
Output Options 4-69
Program Output 4-69
Output to a Text File or Printer 4-71
Output to Screen 4-73

RHC Two Stage
Purpose 4-76
Input Screen 4-76
Format of Input File 4-78
Format of Primary Unit Information 4-79
Input from Text Files 4-80
Input from an Access Database 4-83
Input from Excel Spreadsheets 4-86
Summary Screens 4-89
Output Options 4-90
Program Output 4-90
Output to a Text File or Printer 4-93
Output to Screen 4-96

RHC Three Stage
Purpose 4-98
Input Screen 4-98
Format of Input File 4-100
Format of P.U. / S.U. Information 4-101
Input from Text Files 4-103
Input from an Access Database 4-106
Input from Excel Spreadsheets 4-110

Output Options 4-113
Program Output 4-113
Output to a Text File or Printer 4-116
Output to Screen 4-119

Stratified Cluster
Purpose 4-122
Input Screen 4-122
Format of Input File 4-124
Input from a Text File 4-125
Input from an Access Table 4-126
Input from an Excel Spreadsheet 4-129
Complete or Summary Output 4-130
Output Options 4-130
Program Output 4-131
Output to a Text File or Printer 4-133
Output to Screen 4-134

Stratified Multistage
Purpose 4-136
Input Screen 4-136
Input of Data Values 4-138
Output Options 4-139
Program Output 4-139
Output to a Text File or Printer 4-141
Output to Screen 4-142

Poststratification
Purpose 4-143
Input Screen 4-143
Format of Input File 4-145
Input from Text Files 4-146
Input from an Access Database 4-149
Input from Excel Spreadsheets 4-151
Complete or Summary Output 4-153
Output Options 4-154
Program Output 4-155
Output to a Text File or Printer 4-156
Output to Screen 4-158

Unknown Universe Size
Purpose 4-161

Input Screen 4-161
Specify Input Information 4-162
Sample Used to Estimate the Universe Size 4-163
Sample Used for Variable Estimation 4-163
Output Options 4-164
Program Output 4-164
Output to a Text File or Printer 4-166
Output to Screen 4-166

SAMPLE SIZE DETERMINATION

Overview
Variable Sample Size Determination 5-1
Attribute Sample Size Determination 5-1

Variable Sample Size Determination - Unrestricted Using Reported Amounts
Purpose 5-2
Input Screen 5-2
Probe Sample 5-3
Probe Sample in a Text File 5-3
Probe Sample in an Excel Spreadsheet 5-4
Probe Sample in an Access Database Table 5-6
No Probe Sample Used 5-8
Working with the Full Input Screen 5-9
Confidence Level 5-9
Precision 5-9
Program Output 5-10
Explanation of Output 5-10
Output to a Text File or Printer 5-10
Output to Screen 5-11

Variable Sample Size Determination - Unrestricted Using Estimated Error Rate
Purpose 5-13
Input Screen 5-13
Assumptions 5-14
Working with the Full Input Screen 5-14
Confidence Level 5-14
Precision 5-14
Example 5-15
Program Output 5-16
Explanation of Output 5-16
Output to a Text File or Printer 5-16
Output to Screen 5-18

Variable Sample Size Determination - Stratified

Purpose 5-19
Input Screen 5-19
Sample Size is Known / Unknown 5-20
Working with the Full Input Screen 5-20
Confidence Level 5-20
Precision 5-20
Program Output 5-20
Explanation of Output 5-21
Program Output - Total Sample Size is Unknown 5-22
Output to a Text File or Printer 5-22
Output to Screen 5-24
Program Output - Total Sample Size is Known 5-26
Output to a Text File or Printer 5-26
Output to Screen 5-27

Attribute Sample Size Determination

Purpose 5-29
Input Screen 5-29
Input Values 5-30
Confidence Level 5-30
Anticipated Rate of Occurrence 5-30
Universe Size 5-31
Desired Precision Range 5-31
Program Output 5-31
Explanation of Output 5-31
Output to a Text File or Printer 5-32
Output to Screen 5-33

APPENDICES

Data Limitations for RAT-STATS 2007 . . . . . A-1
Random Number Modules A-1
Attribute Modules A-1
Variable Modules A-2
Sample Size Modules A-3

Numerical Accuracy in RAT-STATS 2007 . . . . . B-1

Troubleshooting for RAT-STATS 2007 . . . . . C-1

INTRODUCTION

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
RAT-STATS 2007
VERSION 2

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's 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 **RATSTATS2007V2.exe**. This program can be installed by double clicking on **SetupRS2007V2.exe** After installing RATSTATS2007V2.exe, the program can be executed by clicking on this file. An easier procedure would be to install a shortcut to RATSTATS2007V2.exe on the windows desktop.

The default installation location is C:\Program Files\RATSTATS2007. During the installation procedure, the user has the option of selecting a different location. After the installation is complete, the folder containing RATSTATS2007V2.exe will also contain User Guide and Companion Manual files in pdf format.

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 using 13 certification programs from the National Bureau of Standards to test for various aspects of randomness. The software passed all 13 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

WHOM DO I CONTACT ABOUT RAT-STATS?

If you have questions about the software or suggestions for improvements, you may contact Office of Inspector General, Office of Public Affairs at paffairs@oig.hhs.gov.

WHY IS IT CALLED RAT-STATS?

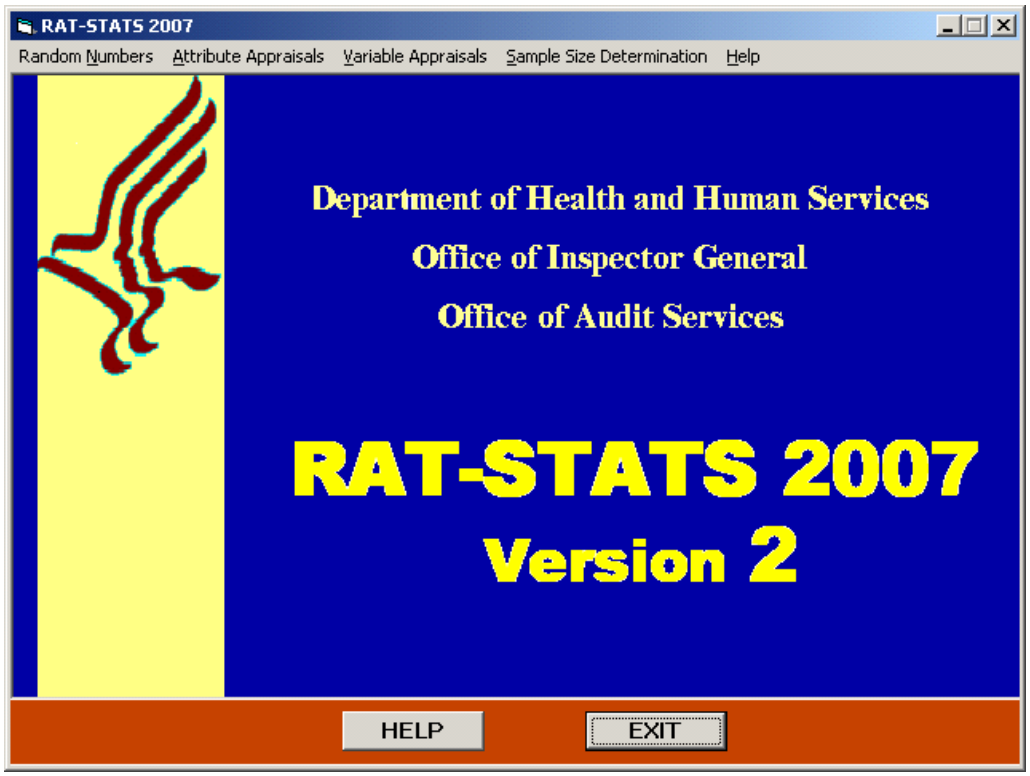
This is the most common question asked about the package. The software was initially developed by the Regional Advanced Techniques Staff (RATS) in San Francisco. After an informal naming contest in the RATS office, the name **RAT-STATS** won by a process of elimination.

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

SINGLE-STAGE NUMBERS

SETS OF TWO NUMBERS

SETS OF THREE NUMBERS

SETS OF FOUR NUMBERS

FRAMES - SINGLE STAGE

FRAMES - SETS OF TWO

RHC SAMPLE SELECTION

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 generator programs.

Note: RAT-STATS has been designed to ignore invalid keystrokes in all the Random Number modules, such as a minus sign or period when entering a universe size.

Attribute Appraisals

UNRESTRICTED

STRATIFIED

TWO-STAGE UNRESTRICTED

THREE-STAGE UNRESTRICTED

RHC TWO-STAGE

RHC THREE-STAGE

STRATIFIED CLUSTER

STRATIFIED MULTISTAGE

Attribute appraisals allow 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 attribute appraisal programs. Section three of this guide explains each of the attribute appraisal programs.

Note: RAT-STATS has been designed to ignore invalid keystrokes in all the Attribute Appraisal modules, such as a minus sign or period when entering a universe size.

Variable Appraisals

UNRESTRICTED

STRATIFIED

TWO-STAGE UNRESTRICTED

THREE STAGE UNRESTRICTED

RHC TWO-STAGE

RHC THREE-STAGE

STRATIFIED CLUSTER

STRATIFIED MULTISTAGE

POST STRATIFICATION

UNKNOWN UNIVERSE SIZE

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 variable appraisal programs. Section four of this guide explains each of the variable appraisal programs.

Note: RAT-STATS has been designed to ignore invalid keystrokes in all the Variable Appraisal modules, such as a minus sign or period when entering a universe size.

Sample Size Determination

VARIABLE SAMPLE SIZE DETERMINATION
- - Unrestricted - Using a Probe Sample
- - Unrestricted - Using Estimated Error Rate
- - Stratified

ATTRIBUTE 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 (Using a Probe Sample) 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 Unrestricted (Using Estimated Error Rate) module determines an optimum sample size by first estimating the mean and standard deviation of the difference amounts using the reported amounts and an estimated error rate. The Variable Stratified module will determine optimum sample sizes for situations where the total sample size is either 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.

By selecting **SAMPLE SIZE DETERMINATION**, the user will see a window appear with the sample size determination programs. Section five of this guide explains each of the sample size determination programs.

Note: RAT-STATS has been designed to ignore invalid keystrokes in all the Sample Size modules, such as a minus sign or period when entering a universe size.

Random Numbers

OVERVIEW

SINGLE-STAGE NUMBERS

SETS OF TWO NUMBERS

SETS OF THREE NUMBERS

SETS OF FOUR NUMBERS

FRAMES - SINGLE STAGE

FRAMES - SETS OF TWO

RHC SAMPLE SELECTION

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 3,000, 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.

Sets of Two Numbers

This module is used when sample items can be easily identified in a two-step process. For example, the user has a computer printout of a universe that contains page numbers and no item numbers. The user could use this module and enter the range of page numbers (e.g., 1 to 150) and then enter the range of the maximum lines on a page (e.g., 1 to 66). This approach might be more expedient to a user than to number all of the items in the listing.

Sets of Three Numbers

This module is used when sample items can be easily identified in a three-step process. For example, the user is sampling from monthly listings of transactions for a 1-year period of time. The first set of numbers entered would be for the range of months (e.g., 1 to 12). The second set of numbers entered would be for the range of the maximum number of pages for one monthly listing (e.g., 1 to 842). The third set of numbers would be for the range of the maximum lines on a page (e.g., 1 to 66).

Sets of Four Numbers

This module would be used in situation similar to the "Sets of Three Numbers" with the additional level of selection. For example, the user might have multiple years to select items from and might use a year, month, page, line approach to draw the sample items.

Frames - Single Stage

The user may find a situation where the universe of items has large gaps in the numbering or the numbering scheme is reset at the beginning of each period. For example, the user may be interested in looking at a universe of 712 checks that are numbered 1,201 through 1,483 and 2,833 through 3,261. The module allows the user to enter in these two range sets (frames). The module will calculate the universe (e.g., 712) and for each random number selected (e.g., 10), determine the set each random selection is in (e.g., 1) and the number within that set (e.g., 1,210).

The user may also have a situation where vouchers are numbered starting with 1 at the beginning of each month. For a 6-month review, the user would have six frames of voucher numbers. The range for each month would be entered and the module would calculate the overall universe of items. The output would display the random number drawn and the month and voucher identified by the random number.

Frames - Sets of Two

Similar to "Frames - Single Stage," the user may have a situation where the universe has gaps or repeats on a periodic basis. In addition, the user may see that a two-step process (e.g., page and line number) is the most efficient way to select the sample items. For example, the user may be working with monthly computer listings. The number of pages may vary significantly from month to month. By using this module, the user could establish the frame (e.g., pages) for each month and then indicate the maximum number of lines on a page (e.g., 66). If the user found that the maximum number of lines changed significantly from month to month, then the user has the option of entering the maximum boundaries (e.g., lines) for each frame (e.g., pages).

RHC Sample Selection

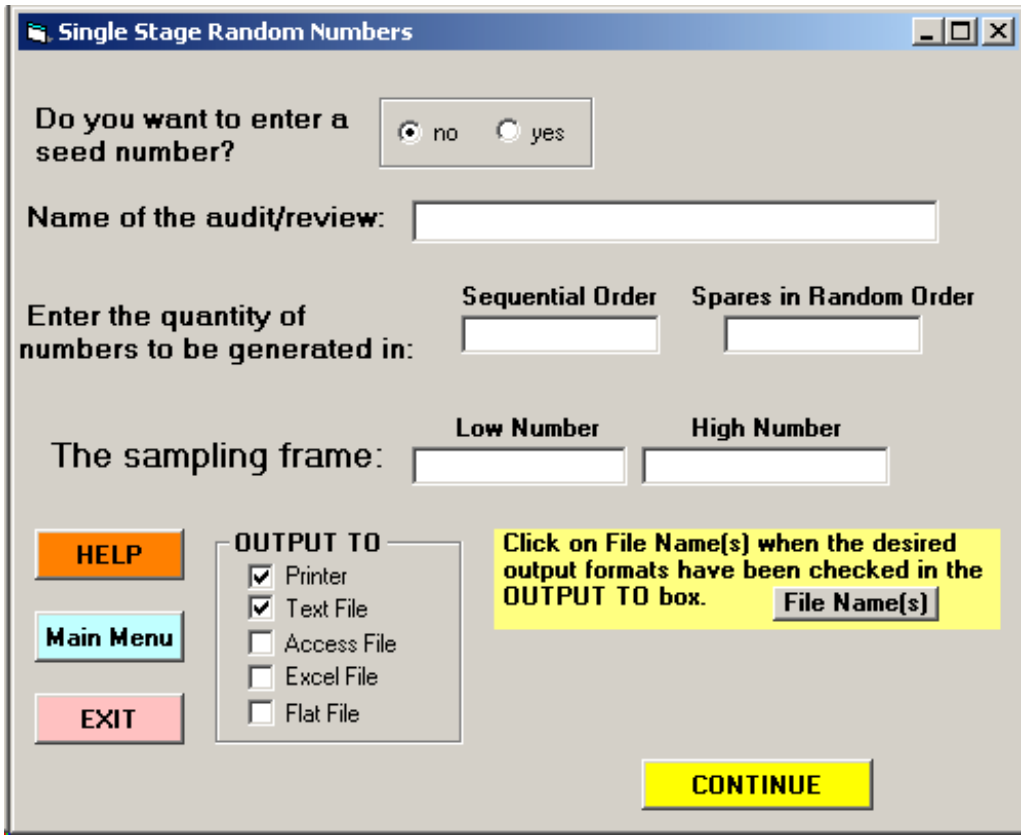
In certain situations a user may want to draw a multistage sample with the probability greater for selecting "larger" units in the universe. For example, a user may want to take an inventory of items at various warehouses. However, the user may want the larger warehouses to have a greater chance for selection. The statistical methodology developed by Rao, Hartley, and Cochran allows the user to weight (e.g., square footage at each warehouse) the primary (and secondary, if it is a three-stage sample) units and thereby increase the chance for larger units to be sampled. This sampling methodology requires that the appropriate appraisal software be used.

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 to 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.

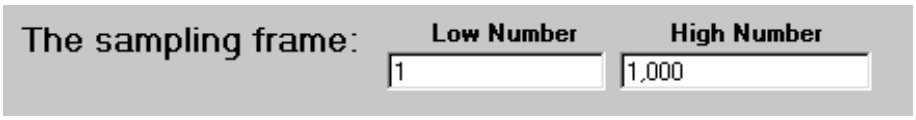
Enter the quantity of numbers to be generated in: Sequential Order

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 arranged 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 this box can be left blank or a "0" (zero) can 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 this box can be left blank or a "0" (zero) can be entered.

Entering the sampling frame

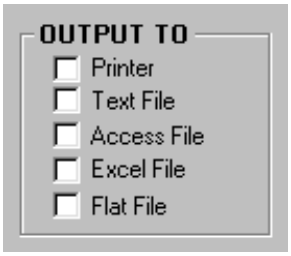


The low and high numbers in the sampling frame are the boundaries of the frame from which the user will be sampling. If the frame is a computer listing numbered 1 through 1,000 then the low entry will be 1 and the high entry will be 1,000. If the frame is a check register with checks numbered between 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 be sent to printer, text file, Access table, Excel spreadsheet, or flat file. The program always concludes with a summary on the screen.

Program Output



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions..

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)**. If the OUTPUT TO **Text File** is selected, the standard Windows "Save" file screen will appear. Type in the name of the file in the **File name** box and click on **Save**. The output file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). After saving the file, the program will return to the original input screen for this module.

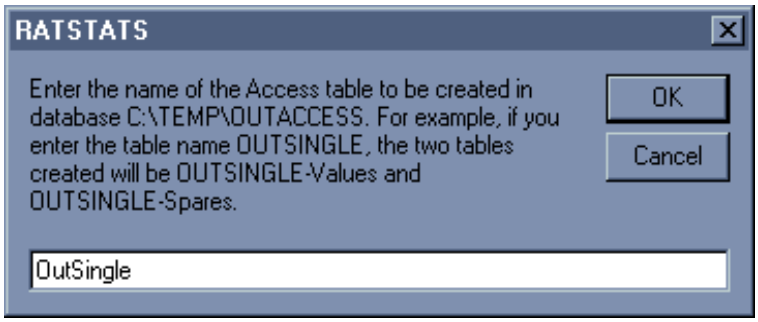
The output shown next is file C:\ TEMP\OUTDISK.TXT and is the result of generating 10 random numbers between 1 and 1,000 along with 4 spares. For each random number generated, two pieces of information are provided. The first is the order of selection. This number indicates the order in which the random 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: 10/12/2004           Random Number Generator          Time: 13:34
                          AUDIT: User Guide Example

SEED NUMBER: 48863.78                            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.

Selection
  Order     Value
    10         22 ◄ The 10 random values start here.
     8        141
     4        236
     3        247
     7        257
     9        301
     2        643
     6        650
     1        718
     5        821
    11        980 ◄ The 4 spare values start here.
    12        507
    13        440
    14        557

SUMMATION OF RANDOM NUMBERS = 6,520
```

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

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Fill in the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 tables OutSingle-Values and OutSingles-Spares, shown next, are the result of generating 10 random numbers between 1 and 1,000 along with 4 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.

The OutSingle-Values table contains ten random numbers, the seed number, the date, and the time.

OutSingle-Values : Table

Order	Value	Seed-Number	Date	Time
10	22	48863.78	9/7/2005	12:19:00 PM
8	141			
4	236			
3	247			
7	257			
9	301			
2	643			
6	650			
1	718			
5	821			

The OutSingle-Spares table contains four spare values.

OutSingle-Spares : Table

Order	Value
11	980
12	507
13	440
14	557

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Fill in the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet 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.

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

	A	B	C	D
1	**Department of Health and Human Services**			
2	**OIG - Office of Audit Services**			
3	**Random Number Generator**			
4	**Date:**	10/12/2004	**Time:**	13:34
5	**Audit:**	User Guide Example		
6	**Order**	**Value**	**Seed Number**	**Frame Size**
7	10	22	48863.78	1,000
8	8	141		
9	4	236		
10	3	247		
11	7	257		
12	9	301		
13	2	643		
14	6	650		
15	1	718		
16	5	821		

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

	A	B
1	**Order**	**Value**
2	11	980
3	12	507
4	13	440
5	14	557

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

NOTE: The user must first exit RAT-STATS in order to view this file.

Output to a Flat File

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. The name of the file for this illustration is C:\TEMP\OUTFLAT.

The following output is the text file created by selecting "Flat File" as one of the output options. This file is the output file created when generating 10 random numbers between 1 and 1,000 along with 4 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.

```
                 Department of Health and Human Services
                     OIG - Office of Audit Services
Date: 10/12/2004         Random Number Generator                Time: 13:34
                          AUDIT: User Guide Example

SEED NUMBER: 48863.78                            FRAME SIZE:           1,000


     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 14


00100000000022 ⬅ The 10 random values start here.
00080000000141
00040000000236
00030000000247
00070000000257
00090000000301
00020000000643
00060000000650
00010000000718
00050000000821
00110000000980 ⬅ The 4 spare values start here.
00120000000507
00130000000440
00140000000557

SUMMATION OF RANDOM NUMBERS = 6,520
```

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

Output to a Printer

If the user selects the printer for output, the standard Windows "Print" dialog box 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 1,000 along with 4 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: 10/12/2004 RANDOM NUMBER GENERATOR TIME: 13:34
AUDIT: User Guide Example

SEED NUMBER: 48863.78 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 -	22	8-	141	4 -	236
3 -	247	7 -	257	9 -	301
2 -	643	6 -	650	1 -	718
5 -	821				

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/12/2004 RANDOM NUMBER GENERATOR TIME: 13:34
AUDIT: User Guide Example

SEED NUMBER: 48863.78 FRAME SIZE: 1,000

4 RANDOM NUMBERS IN GENERATED ORDER

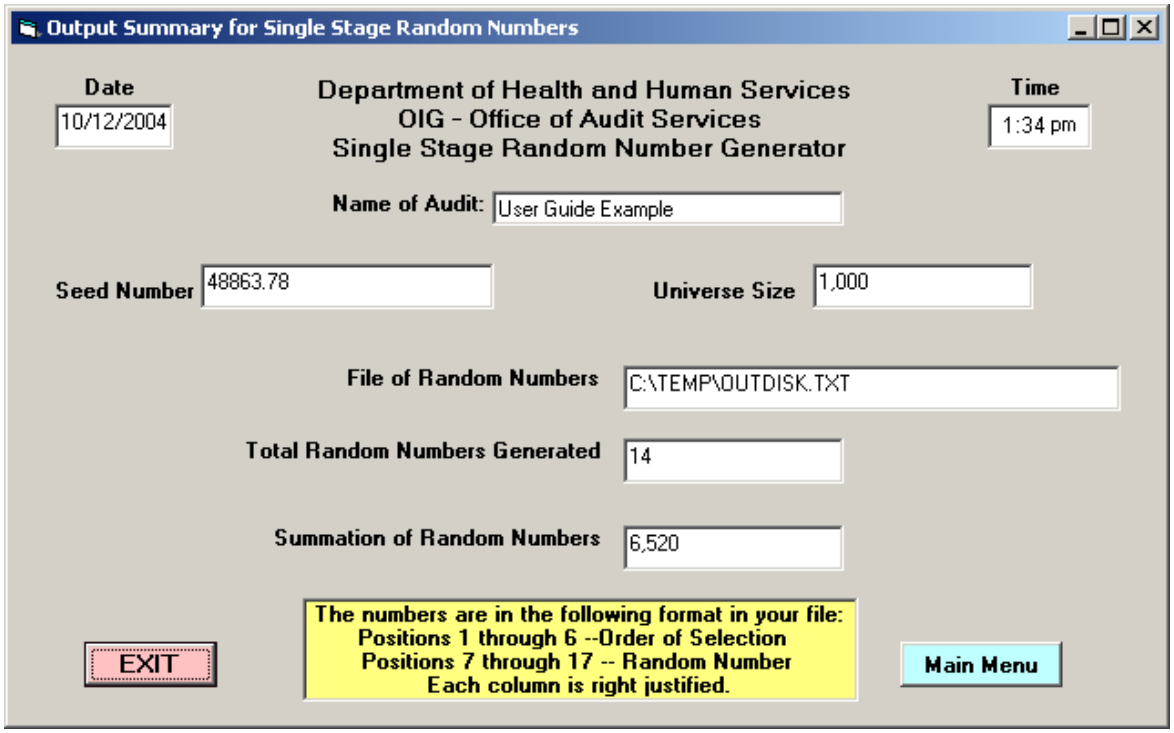
ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER
11 -	980	12 -	507	13 -	440
14 -	557				

SUMMATION OF RANDOM NUMBERS 6,520

Output to Screen

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

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



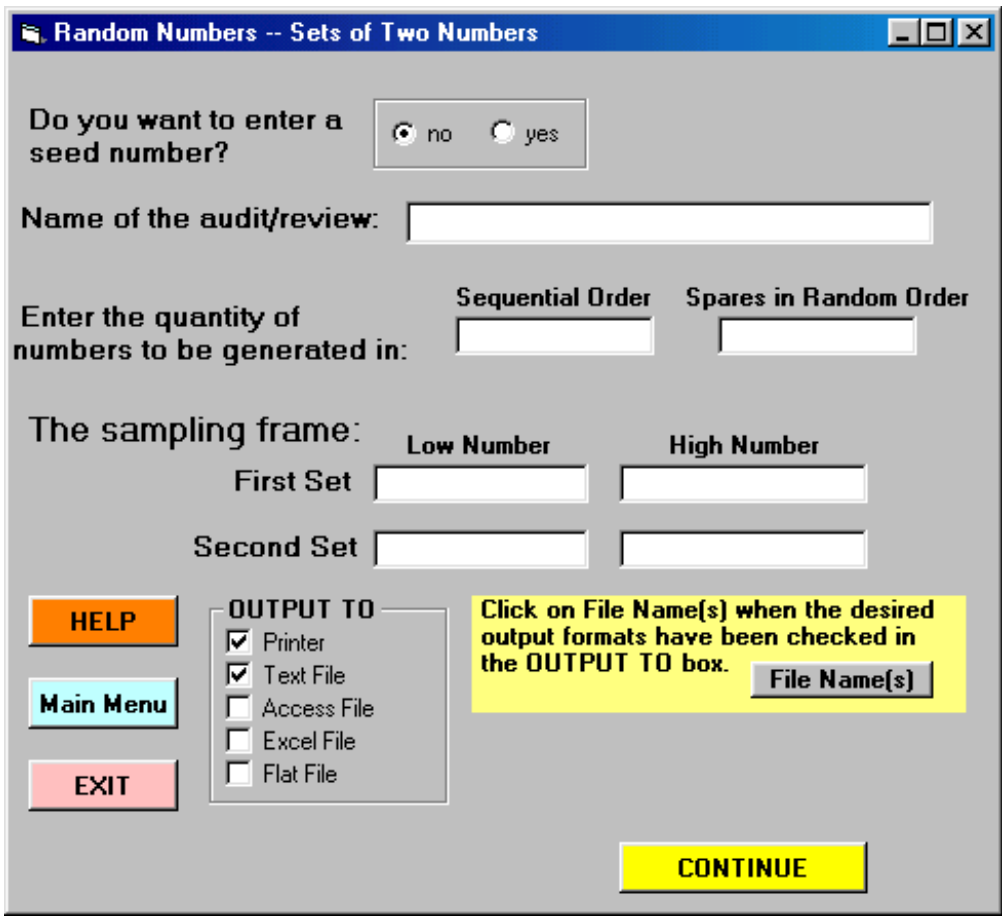
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

SETS OF TWO NUMBERS

Purpose

This program will generate an unduplicated pair (called a set) of random numbers (maximum of 2,000 pairs) for the user. This program may be more efficient for a user when a sample item can be quickly identified through a two-step process (e.g., page and line number).

Input Screen



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.

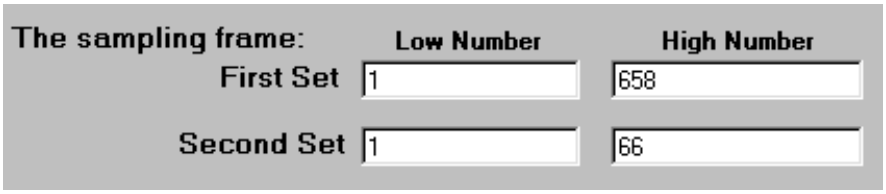
Enter the quantity of numbers to be generated in: sequential order

The quantity of random numbers 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 printed 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 this box can be left blank or a "0" (zero) can 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 printed in the order selected. If the quantity desired is zero, then this box can be left blank or a "0" (zero) can be entered.

Entering the sampling frame



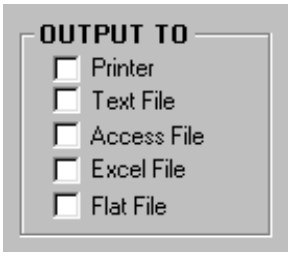
The program prompts the user to enter the low and high numbers of the sampling frame that constitute the first range in the pair of ranges for the entire frame. If, for example, the user was

planning to select items from a computer printout that had pages numbered 1 through 658 and had 66 lines on each page, then the low for the first set (pages) would be 1 and the high for the first set would be 658. The low for the second set (lines on a page) would be 1 and the high for the second set would be 66. The overall frame size for this sample would be 43,428 (658 times 66). For ease of use, the path to the sample items should dictate the selection of frame boundaries that are placed in each set. Thus, for the above example, the user should give the page boundaries for the first set and the line boundaries for the second set. While doing this in reverse order (line boundaries first) is permissible, the sequential ordering of the pairs of numbers would be in ascending order by line instead of by page. Such ordering would normally increase the time required by the user to locate and select the sample items.

Output Options

The program allows for five output options. The user may select the output 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



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

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)**. If the OUTPUT TO **Text File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

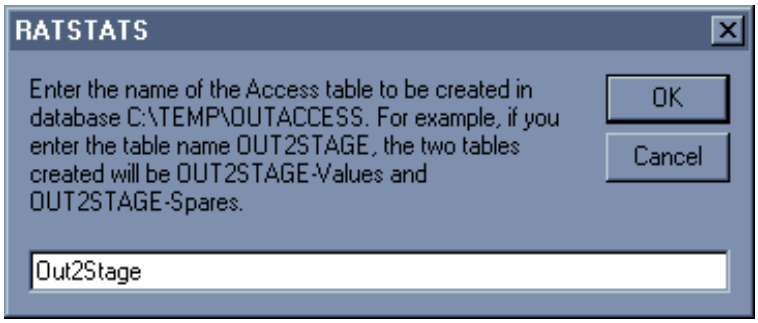
The output shown below is file C:\ TEMP\OUTDISK.TXT and is the output using the values in the preceding **Entering the sampling frame** display. Ten random values were selected along with four spares. For each set of random numbers generated, three pieces of information are provided. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/12/2004             Random Number Generator              Time: 16:57
                           AUDIT: User Guide Example

SEED NUMBER: 61061.36                            FRAME SIZE:          43,428

     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 - FIRST NUMBER OF SET
   POSITIONS 18 THROUGH 30 - SECOND NUMBER OF SET
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.


Selection   First       Second
  Order    Number       Number
     9       147           47  ◄ The 10 random values start here.
     2       165           26
     6       169            6
     4       200           31
     8       250           56
     1       309           41
     5       471           50
     7       491           48
     3       510           29
    10       598           32
    11        48           42  ◄ The 4 spare values start here.
    12       101           56
    13       327            5
    14       288           30
```

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

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 "Out2Stage" will be used. The program will then create two tables (Out2Stage-Values and Out2Stage-Spares) within database C:\TEMP\OUTACCESS.MDB.

The tables Out2Stage-Values and Out2Stage-Spares, shown next, are the result of generating 10 random numbers along with 4 spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, three 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 next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The seed number used by the random number generator is in the far-right column. The Out2Stage-Values table contains 10 random numbers, the seed number, the date, and the time.

Out2Stage-Values : Table

Order	First-Value	Second-Value	Seed-Number	Date	Time
9	147	47	61061.36	9/7/2005	12:26:00 PM
2	165	26			
6	169	6			
4	200	31			
8	250	56			
1	309	41			
5	471	50			
7	491	48			
3	510	29			
10	598	32			

The Out2Stage-Spares table contains four spare values.

Out2Stage-Spares : Table

Order	First-Value	Second-Value
11	48	42
12	101	56
13	327	5
14	288	30

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet C:\TEMP\OUTEXCEL.XLS and is the result of generating 10 random numbers along with 4 spares using the values in the preceding **Entering the sampling frame** display. By clicking on the tab labeled "SPARES," the user will obtain the output immediately following. For each set of random numbers generated,

three 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 next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. 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.

These are the 10 values in the Excel spreadsheet named "VALUES" located in C:\TEMP\OUTEXCEL.XLS.

	A	B	C	D	E
1	**Department of Health and Human Services**				
2	**OIG - Office of Audit Services**				
3	**Random Number Generator**				
4	**Date:**	10/12/2004	**Time:**	16:57	
5	**Audit:**	User Guide Example			
6	**Order**	**First Value**	**Second Value**	**Seed Number**	**Frame Size**
7	9	147	47	61061.36	43,428
8	2	165	26		
9	6	169	6		
10	4	200	31		
11	8	250	56		
12	1	309	41		
13	5	471	50		
14	7	491	48		
15	3	510	29		
16	10	598	32		

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

	A	B	C
1	**Order**	**First Value**	**Second Value**
2	11	48	42
3	12	101	56
4	13	327	5
5	14	288	30

NOTE: The user must first exit RAT-STATS in order to view this file.

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

Output to a Flat File

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file form will appear. Fill in the name of the file in the **File name** box. The name of the file for this illustration is C:\TEMP\OUTFLAT.

The output immediately following is the text file created by selecting "Flat File" as one of the output options. This output is the file created when generating 10 random numbers along with 4 spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, three pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. 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.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/12/2004            Random Number Generator              Time: 16:57
                          AUDIT: User Guide Example

SEED NUMBER: 61061.36                            FRAME SIZE:          43,428

     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 14

000900000001470000000047 ⬅ The 10 random values start here.
000200000001650000000026
000600000001690000000006
000400000002000000000031
000800000002500000000056
000100000003090000000041
000500000004710000000050
000700000004910000000048
000300000005100000000029
001000000005980000000032
001100000000480000000042 ⬅ The 4 spare values start here.
001200000001010000000056
001300000003270000000005
001400000002880000000030
```

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

Output to a Printer

If the user selects the printer for output, the standard Windows "Printer" dialog box will appear. Select the appropriate printer and click on **OK**.

The output immediately following is the printer output when generating 10 random numbers along with 4 spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, three pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. For numbers printed in sequential order, the lowest set of numbers is printed first. The subsequent sets of random numbers are in sequential order going from left to right across the page. Each line will have a maximum of two sets of random numbers with the last line having one or two sets depending on the quantity requested.

The random sets of numbers in generated order will begin on a new page after all the sets in sequential order have been printed.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/12/2004 RANDOM NUMBER GENERATOR TIME: 16:57
AUDIT: User Guide Example

SEED NUMBER: 61061.36 FRAME SIZE: 43,428

10 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	ORDER OF SELECTION	FIRST NBR.	SECOND NBR.
9 -	147	47	2 -	165	26
6 -	169	6	4 -	200	31
8 -	250	56	1 -	309	41
5 -	471	50	7 -	491	48
3 -	510	29	10 -	598	32

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/12/2004 RANDOM NUMBER GENERATOR TIME: 16:57
AUDIT: User Guide Example

SEED NUMBER: 61061.36 FRAME SIZE: 43,428

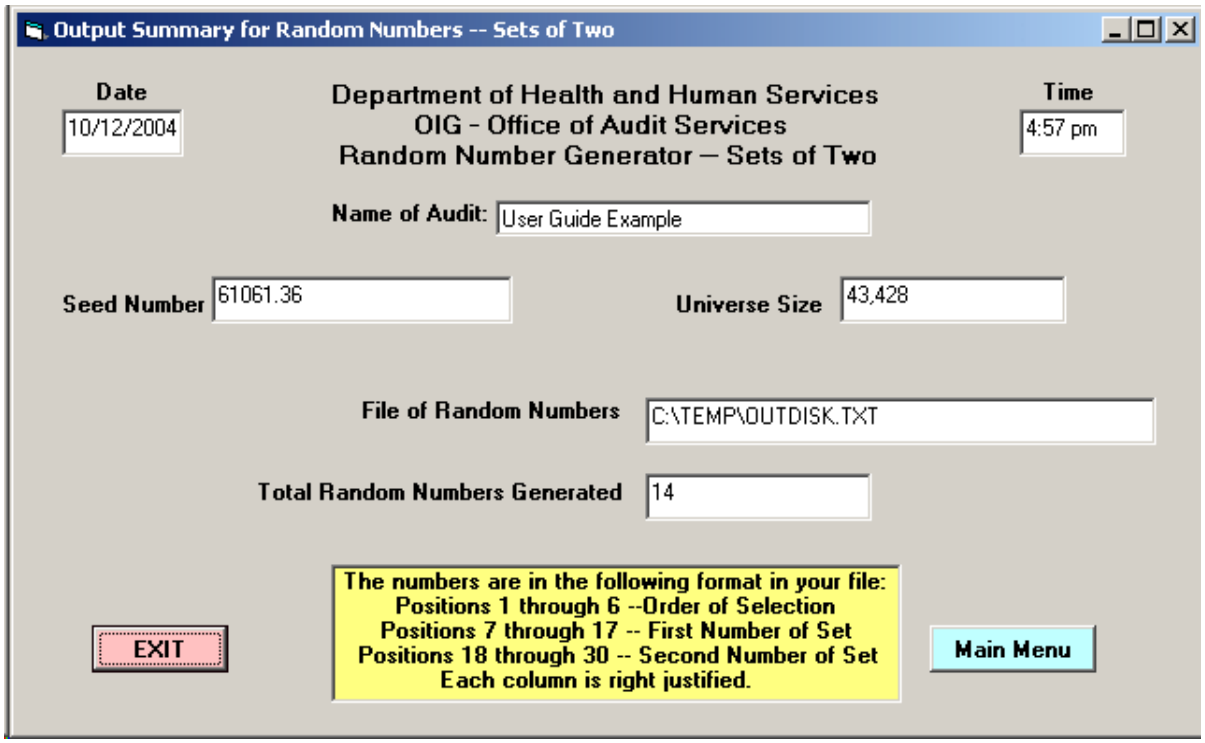
4 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	ORDER OF SELECTION	FIRST NBR.	SECOND NBR.
11 -	48	42	12 -	101	56
13 -	327	5	14 -	288	30

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

Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



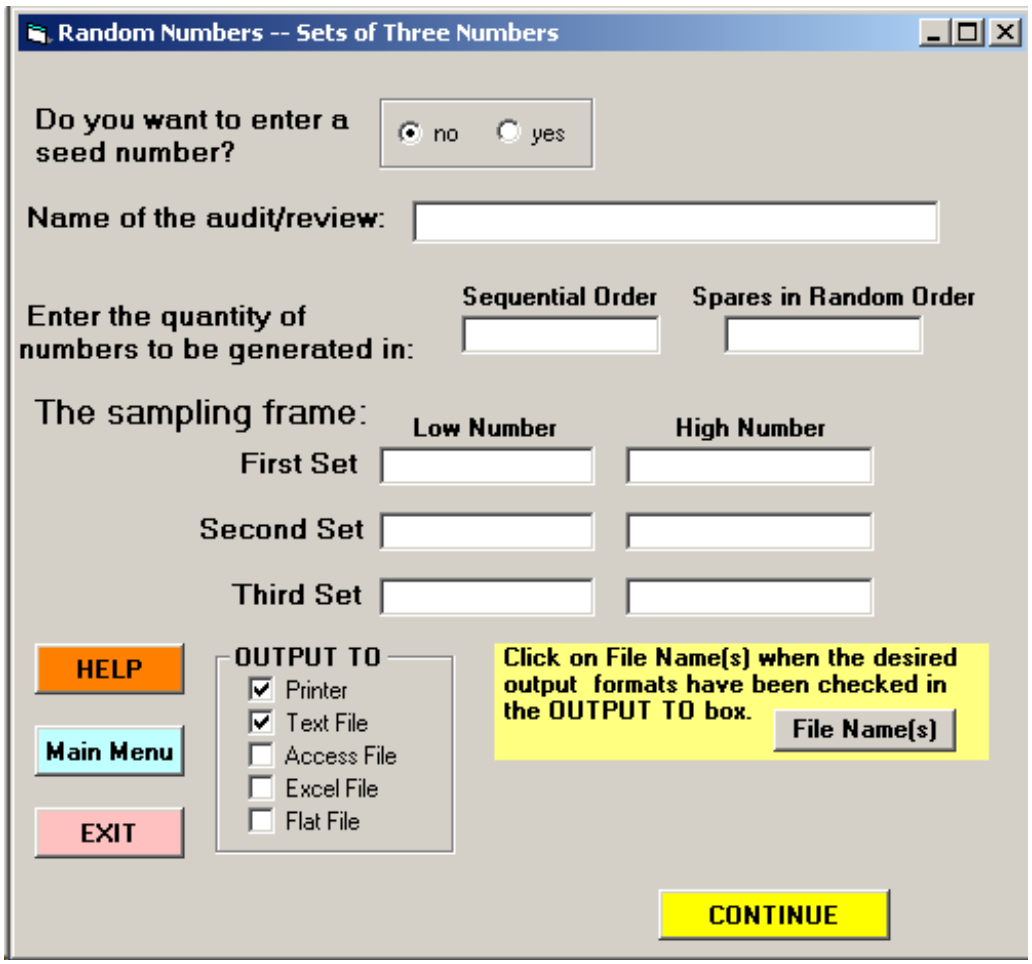
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

SETS OF THREE NUMBERS

Purpose

This program will generate an unduplicated set of random numbers (maximum of 2,000 sets) for the user. This program may be more efficient for a user when a sample item can be quickly identified through a three-step process (e.g., month, page and line number).

Input Screen



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.

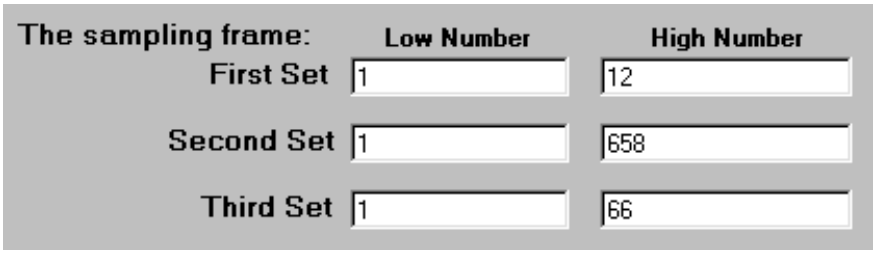
Enter the quantity of numbers to be generated in: sequential order

The quantity of random numbers 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 printed 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 this box can be left blank or a "0" (zero) can 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 printed in the order selected. If the quantity desired is zero, then this box can be left blank or a "0" (zero) can be entered.

Entering the sampling frame



The low and high numbers in the frame are the boundaries of the frame from which the user will

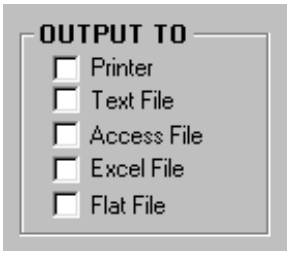
be sampling. If, for example, the user was planning to select items from a year's worth of computer printouts that had pages numbered 1 through 658 each month and had 66 lines on each page, then the first set would be months; the second set would be pages; and the third set would be lines. For the first set the low would be 1 and the high would be 12. The low for the second set would be 1 and the high would be 658. The low for the third set would be 1 and the high would be 66. The overall frame size for this sample would be 521,136 (12 times 658 times 66).

For ease of use, the path to the sample items should dictate the order of frame boundaries that are placed in each set. Thus, for the above example, the user should give the month boundaries for the first set, page boundaries for the second set and the line boundaries for the third set. While doing this in reverse order (line boundaries first) is permissible, the sequential ordering of the sets of numbers would be in ascending order by line instead of by page or month. Such ordering would normally increase the time required by the user to locate and select the sample items.

Output Options

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

Program Output



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

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)**. If the OUTPUT TO **Text File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

The output shown below is file C:\ TEMP\OUTDISK.TXT and is the output using the values in the preceding **Entering the sampling frame** display. Four random values were selected along with two spares. For each set of random numbers generated, four 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 next piece of information is the random number generated from the first set of frame boundaries. The third piece of information is the random number generated from the second set of boundaries. The fourth piece of information is the random number generated from the third set of boundaries.

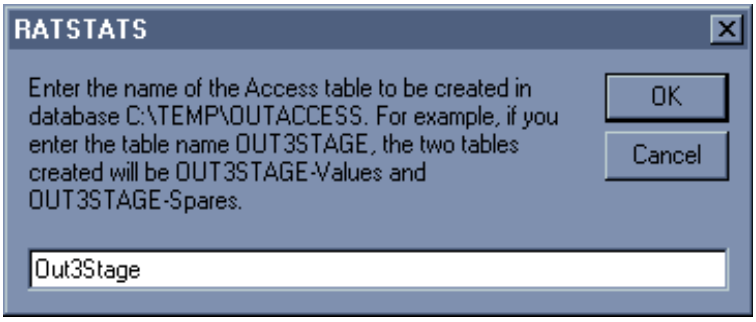
```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/13/2004            Random Number Generator                Time: 9:23
                          AUDIT: User Guide Example

SEED NUMBER: 33818.06                              FRAME SIZE:         521,136

     FILE OF RANDOM NUMBERS: C:\TEMP\OUTDISK.TXT

     TOTAL RANDOM NUMBERS GENERATED: 6

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR
FILE:
   POSITIONS 1 THROUGH  6 - ORDER OF SELECTION
   POSITIONS 7 THROUGH 17 - FIRST NUMBER OF SET
   POSITIONS 18 THROUGH 30 - SECOND NUMBER OF SET
   POSITIONS 31 THROUGH 43 - THIRD NUMBER OF SET
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.
```

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

Selection Order	First Number	Second Number	Third Number	
1	3	484	54	⬅ **The 4 random values start here**
2	9	606	27	
3	10	192	19	
4	12	513	59	
5	6	574	56	⬅ **The 2 spare values start here**
6	2	615	52	

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 for this illustration). The table name "Out3Stage" will be used. The program will then create two tables (Out3Stage-Values and Out3Stage-Spares) within database C:\TEMP\OUTACCESS.

The table Out3Stage-Values shown below is the result of generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, four 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 next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth piece of information is the random number generated from the third set of boundaries. The seed number used by the random number generator is in the far-right column. The Out3Stage-Values table contains four random numbers, the seed number, the date, and the time.

Out3Stage-Values : Table

Order	First-Value	Second-Value	Third-Value	Seed-Number	Date	Time
1	3	484	54	33818.06	9/7/2005	12:31:00 PM
2	9	606	27			
3	10	192	19			
4	12	513	59			

The Out3Stage-Spares table containing the two spare values is shown below.

Out3Stage-Spares : Table

Order	First-Value	Second-Value	Third-Value
5	6	574	56
6	2	615	52

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet C:\TEMP\OUTEXCEL.XLS and is the result of generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. By clicking on the tab labeled "SPARES," the user will obtain the spreadsheet containing the spare values. For each set of random numbers generated, four pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries. 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.

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

	A	B	C	D	E	F
1	**Department of Health and Human Services**					
2	**OIG - Office of Audit Services**					
3	**Random Number Generator**					
4	**Date:**	10/13/2004	**Time:**	9:23		
5	**Audit:**	User Guide Example				
6	**Order**	**First Value**	**Second Value**	**Third Value**	**Seed Number**	**Frame Size**
7	1	3	484	54	33818.06	521,136
8	2	9	606	27		
9	3	10	192	19		
10	4	12	513	59		

	A	B	C	D
1	**Order**	**First Value**	**Second Value**	**Third Value**
2	5	6	574	56
3	6	2	615	52

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

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

Output to a Flat File

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file form will appear. Fill in the name of the file in the **File name** box. The name of the file for this illustration is C:\TEMP\OUTFLAT.

The output immediately following is the text file created by selecting "Flat File" as one of the output options. This output is the file created when generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, four pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries. 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.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/13/2004            Random Number Generator              Time: 9:23
                          AUDIT: User Guide Example

SEED NUMBER: 33818.06                            FRAME SIZE:         521,136


     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 6

000100000000030000000484000000054  ⬅ The 4 random values start here
000200000000090000000606000000027
000300000000100000000192000000019
000400000000120000000513000000059
000500000000060000000574000000056  ⬅ The 2 spare values start here
000600000000020000000615000000052
```

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

Output to a Printer

If the user selects the printer for output, the standard Windows "Printer" dialog box will appear. Select the appropriate printer and click on **OK**.

The output immediately following is the printer output when generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, four pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries.

The random sets of numbers in generated order will begin on a new page after all the sets in sequential order have been printed.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/13/2004 RANDOM NUMBER GENERATOR TIME: 9:23
AUDIT: User Guide Example

SEED NUMBER: 33818.06 FRAME SIZE: 521,136

4 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.
1 -	3	484	54
2 -	9	606	27
3 -	10	192	19
4 -	12	513	59

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/13/2004 RANDOM NUMBER GENERATOR TIME: 9:23
AUDIT: User Guide Example

SEED NUMBER: 33818.06 FRAME SIZE: 521,136

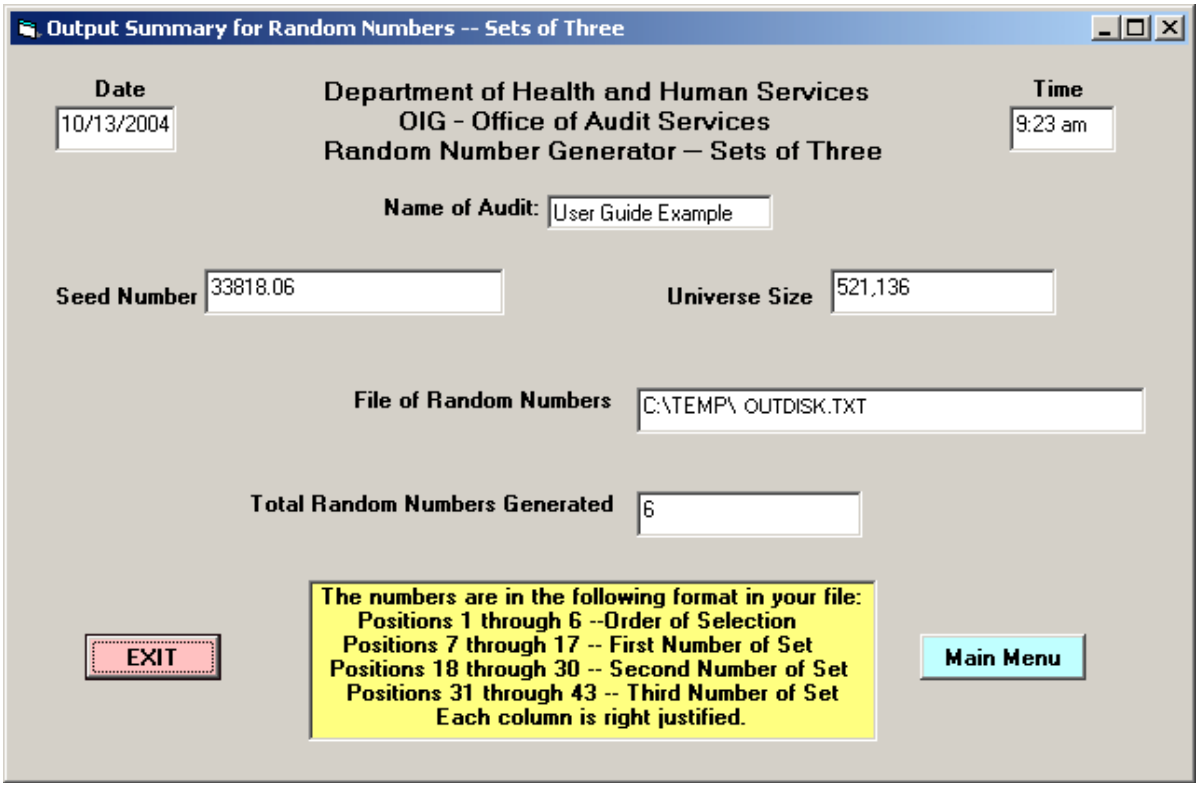
2 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.
5 -	6	574	56
6 -	2	615	52

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

Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



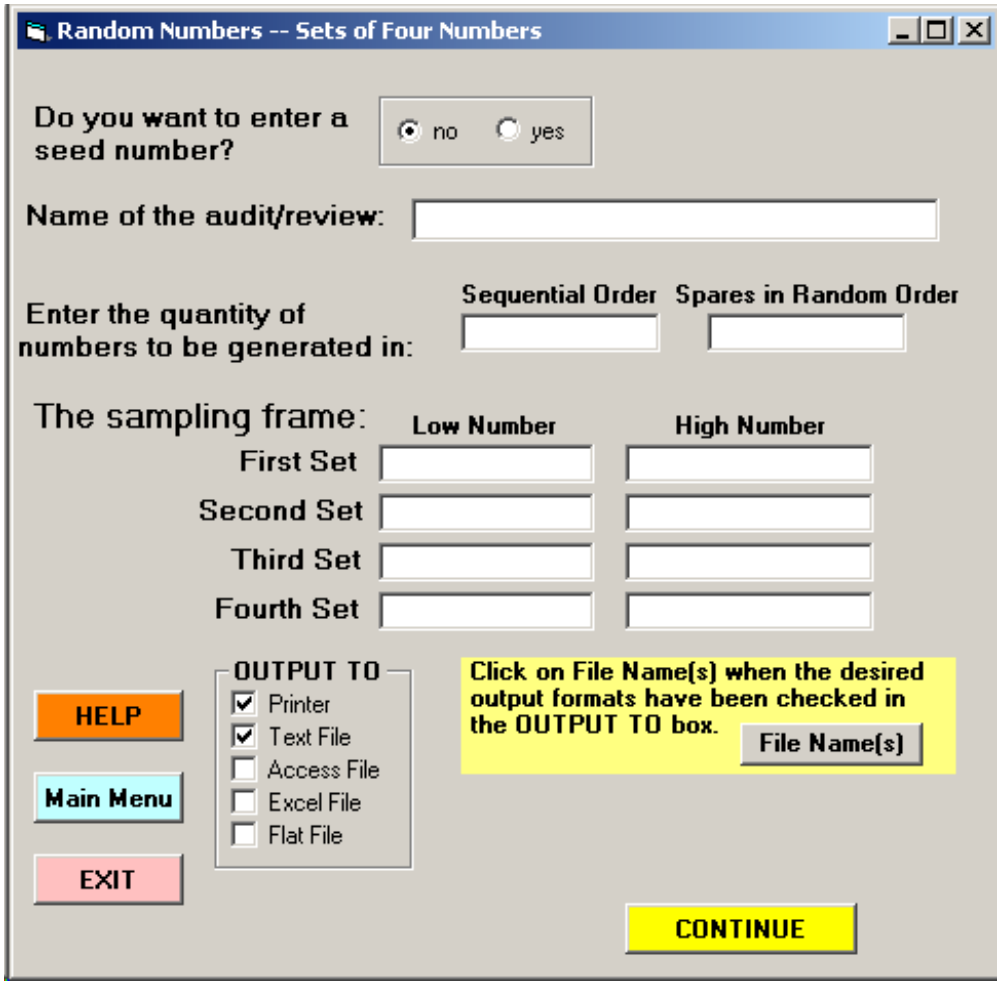
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

SETS OF FOUR NUMBERS

Purpose

This program will generate an unduplicated set of random numbers (maximum of 2,000 sets) for the user. This program may be more efficient for a user when a sample item can be quickly identified through a four-step process (e.g., year, month, page, and line number).

Input Screen



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.

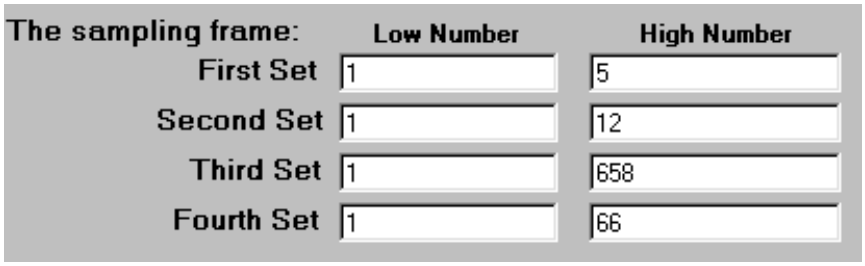
Enter the quantity of numbers to be generated in: sequential order

The quantity of random numbers 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 printed 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 this box can be left blank or a "0" (zero) can 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 printed in the order selected. If the quantity desired is zero, then this box can be left blank or a "0" (zero) can be entered.

Entering the sampling frame



The program prompts the user to enter the low and high numbers for each of the four sets of frames. If, for example, the user was planning to select items from five years' worth of computer printouts that had pages numbered 1 through 658 each month and had 66 lines on each page, then the first set would be year, the second set would be month, the third set would be pages, and the fourth set would be lines. For the first set the low number would be 1 and the high would be 5. The low for the second set would be 1 and the high would be 12. The low for the third set would be 1 and the high would be 658. The low for the fourth set would be 1 and the high would be 66. The overall frame size for this sample would be 2,605,680 (5 times 12 times 658 times 66).

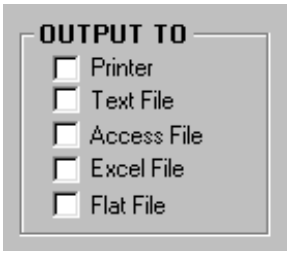
For ease of use, the path to the sample items should dictate the order of frame boundaries that are placed in each set. Thus, for the above example, the user should give the year boundaries for the first set, month boundaries for the second set, page boundaries for the third set, and line boundaries for the fourth set. While doing this in reverse order (line boundaries first) is permissible, the sequential ordering of the sets of numbers would be in ascending order by line instead of by page, month, or year. Such ordering would normally increase the time required by the user to locate and select the sample items.

Output Options

The program allows for five output options. The user may select the output to be sent to printer, text file, Access database, 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 message on the next page will appear:



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

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)**. If the OUTPUT TO **Text File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

The output shown below is file C:\ TEMP\OUTDISK and is the output using the values in the preceding **Entering the sampling frame** display. Three random values were selected along with two spares. For each set of random numbers generated, five 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 next piece of information is the random number generated from the first set of frame boundaries. The third piece of information is the random number generated from the second set of boundaries. The fourth piece of information is the random number generated from the third set of boundaries. The fifth piece of information is the random number generated from the fourth set of boundaries.

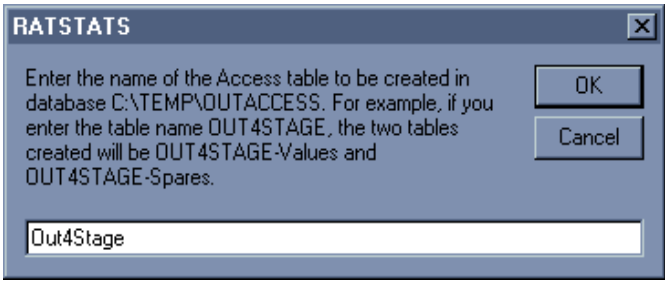
```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/13/2004            Random Number Generator                Time: 9:55
                           AUDIT: User Guide Example

SEED NUMBER: 35731.75                              FRAME SIZE:       2,605,680

     FILE OF RANDOM NUMBERS: C:\TEMP\OUTDISK.TXT

     TOTAL RANDOM NUMBERS GENERATED: 5

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:
   POSITIONS 1 THROUGH  6 - ORDER OF SELECTION
   POSITIONS 7 THROUGH 17 - FIRST NUMBER OF SET
   POSITIONS 18 THROUGH 30 - SECOND NUMBER OF SET
   POSITIONS 31 THROUGH 43 - THIRD NUMBER OF SET
   POSITIONS 44 THROUGH 56 - FOURTH NUMBER OF SET
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.
```

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

Selection Order	First Number	Second Number	Third Number	Fourth Number	
1	1	4	114	24	⬅ **The 3 random values**
3	4	10	442	48	**start here**
2	4	10	501	42	
4	4	5	351	55	⬅ **The 2 spare values**
5	4	9	589	20	**start here**

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 for this illustration). The table name "Out4Stage" will be used. The program will then create two tables (Out4Stage-Values and Out4Stage-Spares) within database C:\TEMP\OUTACCESS.

The table Out4Stage-Values shown below is the result of generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, five 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 next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth piece of information is the random number generated from the third set of boundaries. The fifth piece of information is the random number generated from the fourth set of boundaries. The seed number used by the random number generator is in the far-right column. The Out4Stage-Values table contains three random numbers, the seed number, the date, and the time.

Out4Stage-Values : Table

Order	First-Value	Second-Value	Third-Value	Fourth-Value	Seed-Number	Date	Time
1	1	4	114	24	35731.75	9/7/2005	12:39:00 PM
3	4	10	442	48			
2	4	10	501	42			

The table Out4Stage-Spares containing the two spare values is shown below:

Out4Stage-Spares : Table

Order	First-Value	Second-Value	Third-Value	Fourth-Value
4	4	5	351	55
5	4	9	589	20

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet C:\TEMP\OUTEXCEL and is the result of generating three random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. By clicking on the tab labeled "SPARES," the user will obtain the spreadsheet containing the spare values. For each set of random numbers generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries. The fifth number is the random number generated from the fourth set of boundaries. 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.

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

	A	B	C	D	E	F	G
1	**Department of Health and Human Services**						
2	**OIG - Office of Audit Services**						
3	**Random Number Generator**						
4	**Date:**	10/13/2004	**Time:**	9:55			
5	**Audit:**	User Guide Example					
6	**Order**	**First Value**	**Second Value**	**Third Value**	**Fourth Value**	**Seed Number**	**Frame Size**
7	1	1	4	114	24	35731.75	2,605,680
8	3	4	10	442	48		
9	2	4	10	501	42		

The two spares values in the Excel spreadsheet named "SPARES" are shown below:

	A	B	C	D	E
1	**Order**	**First Value**	**Second Value**	**Third Value**	**Fourth Value**
2	4	4	5	351	55
3	5	4	9	589	20

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

Output to a Flat File

When all the desired output formats have been selected in the OUTPUT TO box, click on the button labeled **File Name(s)**. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file form will appear. Fill in the name of the file in the **File name** box. The name of the file for this illustration is C:\TEMP\OUTFLAT.

The output immediately following is the text file created by selecting "Flat File" as one of the output options. This output is the file created when generating three random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries. The fifth number is the random number generated from the fourth set of boundaries. 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.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/13/2004             Random Number Generator                Time: 9:55
                           AUDIT: User Guide Example

SEED NUMBER: 35731.75                            FRAME SIZE:       2,605,680


     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 5

0001000000000100000000040000000114000000002 4  ⬅ The 3 random values start here
0003000000000400000000010000000044200000000 48
0002000000000400000000010000000050100000000 42
0004000000000400000000050000000035100000000 55  ⬅ The 2 spare values start here
0005000000000400000000090000000058900000000 20
```

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

Output to a Printer

If the user selects the printer for output, the standard Windows "Printer" dialog box will appear. Select the appropriate printer and click on **OK**.

The output immediately following is the printer output when generating four random numbers along with two spares using the values in the preceding **Entering the sampling frame** display. For each set of random numbers generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated from the first set of frame boundaries. The third number is the random number generated from the second set of boundaries. The fourth number is the random number generated from the third set of boundaries. The fifth number is the random number generated from the fourth set of boundaries.

The random sets of numbers in generated order will begin on a new page after all the sets in sequential order have been printed.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/13/2004 RANDOM NUMBER GENERATOR TIME: 9:55
AUDIT: User Guide Example

SEED NUMBER: 35731.75 FRAME SIZE: 2,605,680

3 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	FOURTH NBR.
1 -	1	4	114	24
3 -	4	10	442	48
2 -	4	10	501	42

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/13/2004 RANDOM NUMBER GENERATOR TIME: 9:55
AUDIT: User Guide Example

SEED NUMBER: 35731.75 FRAME SIZE: 2,605,680

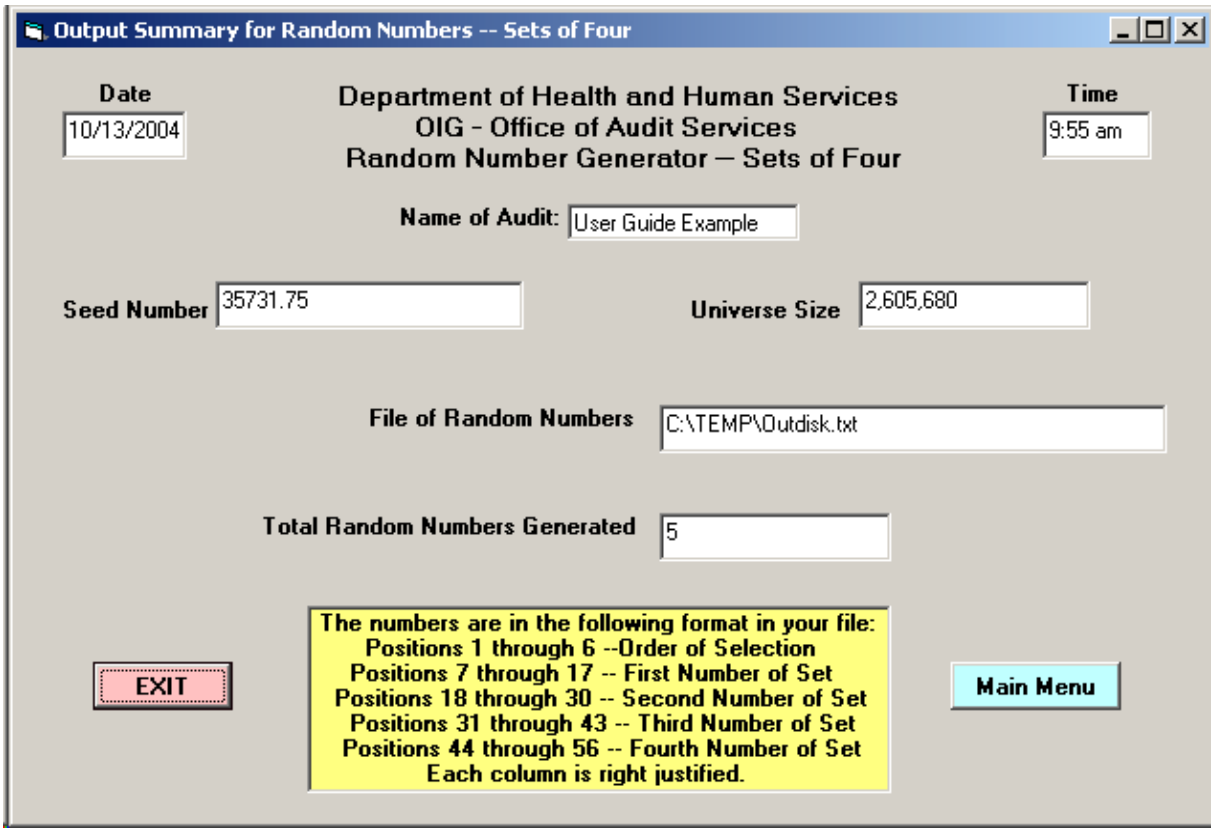
2 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	FOURTH NBR.
4 -	4	5	351	55
5 -	4	9	589	20

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

Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



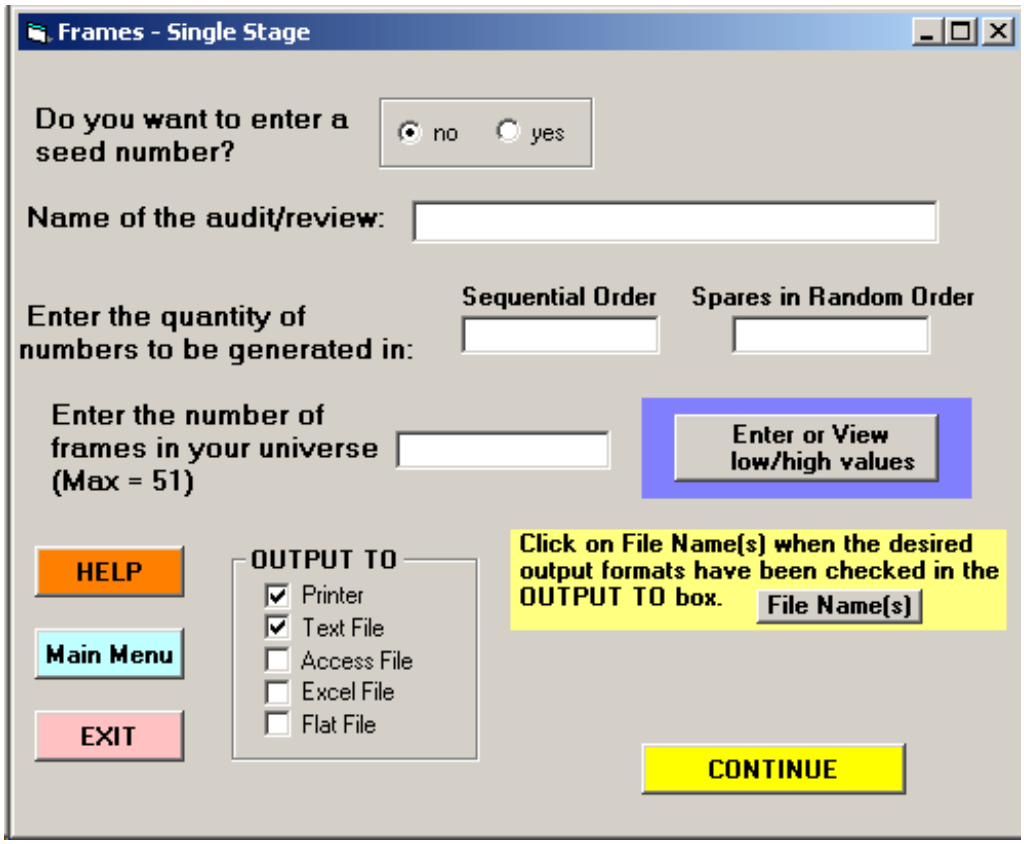
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

FRAMES - SINGLE STAGE

Purpose

This program will generate an unduplicated set of random numbers (maximum of 2,000 sets) for the user. This program is more efficient for a user when the universe of sample items may either contain gaps of numbers or the numbering system repeats within the universe. For example, the universe of transactions that is of interest to the user may be grouped with other transactions in a computer printout that has all items numbered. The transactions of interest may be numbered from 1 through 1,050 and begin again at 8,405 and run through 9,565. Rather than selecting random numbers between 1 and 9,565 and not be able to use the numbers between 1,050 and 8,405 or having the user renumber the subuniverse of desired transactions, this program could be used with the existing numbering system.

Input Screen



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 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 this box can be left blank or a "0" (zero) can 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 this box can be left blank or a "0" (zero) can be entered.

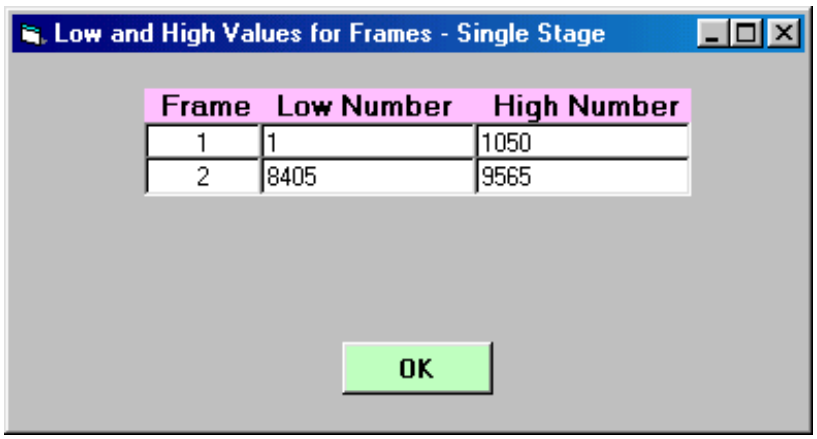
Enter the number of frames in your universe (Max = 51)

A frame is a grouping of contiguous sample items within the universe. In the illustration to follow, the user would have two frames. The first frame would contain items 1 through 1,050 and the second frame would contain items 8,405 through 9,565.

Enter or View low/high values

Based on the frame count entered by the user, the low and high boundaries for each frame must be entered. By clicking on the "Enter or view low/high values" button the following form

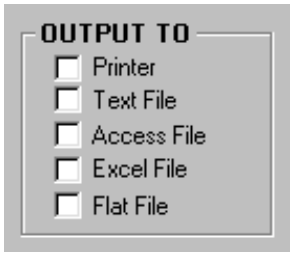
appears if the number of frames is entered as "2." In general, there will be one line in this table for each frame in the universe.



Output Options

The program allows for five output options. The user may select the output 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



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

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)**. If the OUTPUT TO **Text File** is selected, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

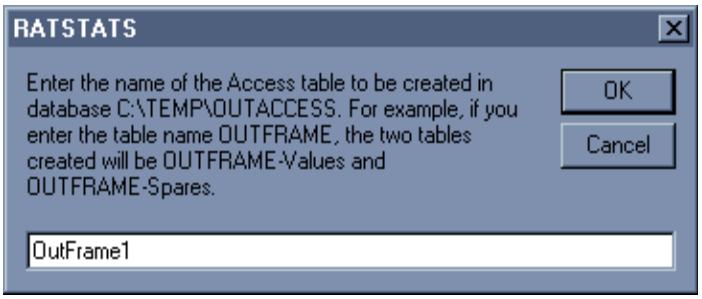
The output shown below is file C:\TEMP\OUTDISK.TXT and is the result of generating three random numbers along with two spares using the low/high values in the earlier screen. For each random number generated, four pieces of information are displayed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/14/2004            Random Number Generator                Time: 9:32
                          AUDIT: User Guide Example

SEED NUMBER: 34346.94                              FRAME SIZE:          2,211

     FILE OF RANDOM NUMBERS: C:\TEMP\OUTDISK.TXT

     TOTAL RANDOM NUMBERS GENERATED: 5

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:
   POSITIONS  1 THROUGH  6 - ORDER OF SELECTION
   POSITIONS  7 THROUGH 20 - RANDOM NUMBER
   POSITIONS 21 THROUGH 29 - FRAME NUMBER
   POSITIONS 30 THROUGH 42 - ITEM NUMBER WITHIN FRAME
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.


Selection     Random      Frame        Item
  Order       Number     Number       Number
     3           119        1          119  ⬅ The 3 random values start here.
     1          1308        2         8662
     2          1603        2         8957
     4           322        1          322  ⬅ The 2 spare values start here.
     5          1553        2         8907
```

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

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 "OutFrame1" will be used. The program will then create two tables (OutFrame1-Values and OutFrame1-Spares) within database C:\TEMP\OUTACCESS.MDB.

The tables OutFrame1-Values and OutFrame1-Spares, shown next, are the result of generating three random numbers along with two spares using the low/high values in the earlier screen. For each random number generated, four 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 next piece of information is the random number generated. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame. The seed number used by the random number generator is in the far-right column.

The OutFrame1-Values table contains three random numbers, the seed number, the date, and the time.

OutFrame1-Values : Table

Order	Number	Frame	Item	Seed-Number	Date	Time
3	119	1	119	34346.94	9/7/2005	12:43:00 PM
1	1308	2	8662			
2	1603	2	8957			

The OutFrame1-Spares table contains two spare values.

OutFrame1-Spares : Table

Order	Number	Frame	Item
4	322	1	322
5	1553	2	8907

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet named "VALUES" located in C:\TEMP\OUTEXCEL.XLS and is the result of generating three random numbers along with two spares using the low/high values in the earlier screen. By clicking on the tab labeled "SPARES" the user will obtain the spreadsheet containing the spare values.

For each random number generated, four 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 next piece of information is the random number generated. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame. The output also contains the program execution date and time, the name of the audit/review, the seed number, and the total size of the sampling frame.

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

	A	B	C	D	E	F
1	**Department of Health and Human Services**					
2	**OIG - Office of Audit Services**					
3	**Random Number Generator**					
4	**Date:**	9/7/2005	**Time:**	12:52		
5	**Audit:**	User Guide Example				
6	**Order**	**Number**	**Frame**	**Item**	**Seed Number**	**Frame Size**
7	3	119	1	119	34346.94	2,211
8	1	1308	2	8662		
9	2	1603	2	8957		

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

	A	B	C	D
1	**Order**	**Number**	**Frame**	**Item**
2	4	322	1	322
3	5	1553	2	8907

NOTE: The user must first exit RAT-STATS in order to view this file.

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

Output to a Flat File

The output immediately following is the text file created by selecting "Flat File" as one of the output options. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file screen will appear. 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 when generating three random numbers along with two spares using the low/high values from the earlier screen. 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.

For each random number generated, four 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 next piece of information is the random number generated. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 9/7/2005             Random Number Generator              Time: 12:52
                           AUDIT: User Guide Example

SEED NUMBER: 34346.94                                FRAME SIZE:        2,211

     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 5

00030000000119000100000000119 ⬅ The three random values start here.
00010000001308000200000008662
00020000001603000200000008957
00040000000322000100000000322 ⬅ The two spare values start here.
00050000001553000200000008907
```

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

Output to a Printer

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

The following printer output is the result of generating three random numbers along with two spares using the low/high values in the earlier screen. For each random number generated, four 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 next piece of information is the random number generated. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame. Summary information will be displayed on the screen.

For numbers printed in sequential order of the printout, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line. The random numbers in generated order will begin on a new page after all the sequential sets have been printed.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/14/2004 RANDOM NUMBER GENERATOR TIME: 9:32
AUDIT: User Guide Example

SEED NUMBER: 34346.94 FRAME SIZE: 2,211

3 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	RANDOM NUMBER	-CONVERTED NUMBER- FRAME	ITEM NUMBER
3 -	119	1	119
1 -	1,308	2	8,662
2 -	1,603	2	8,957

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/14/2004 RANDOM NUMBER GENERATOR TIME: 9:32
AUDIT: User Guide Example

SEED NUMBER: 34346.94 FRAME SIZE: 2,211

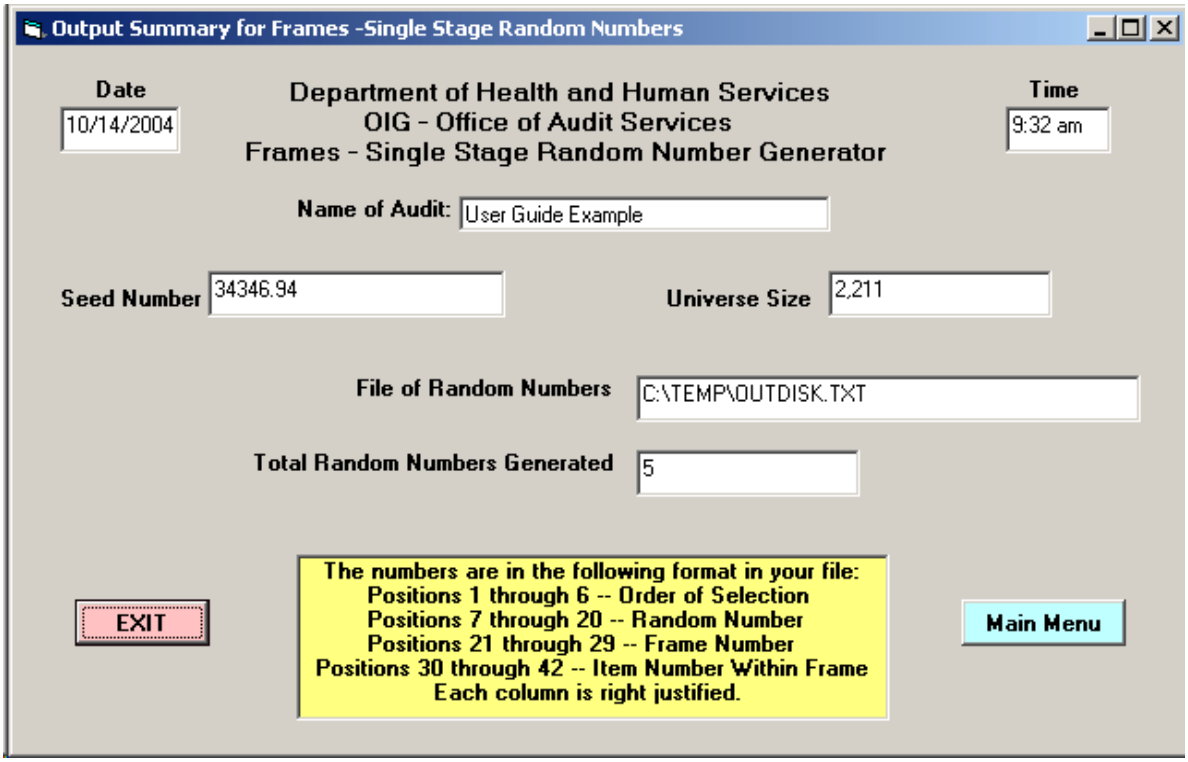
2 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	RANDOM NUMBER	-CONVERTED NUMBER- FRAME	ITEM NUMBER
4 -	322	1	322
5 -	1,553	2	8,907

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

Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



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

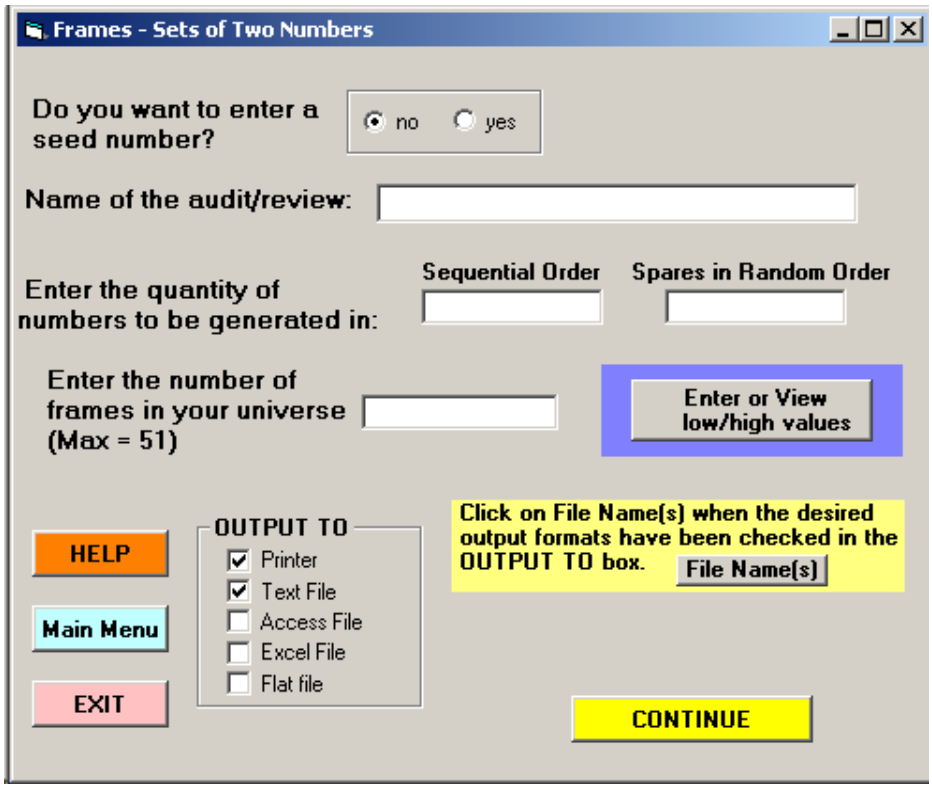
FRAMES - SETS OF TWO

Purpose

This program will generate an unduplicated set of random numbers (maximum of 2,000 sets) for the user. This program may be of use when pairs of random numbers are the most efficient way to locate a sample item and the numbering of the universe of items either has gaps or the numbering system repeats within the universe. For example, the universe of transactions that is of interest to the user may be grouped with other transactions in several computer printouts that have page and line numbers. The transactions of interest may be on pages numbered from 1 through 100 in one volume, pages 1 through 456 in a second volume, and pages 45 through 832 in a third volume. This program could be used with the existing numbering system to select page and line numbers.

Within each frame there may be a consistent number of items, such as 66 lines per page, or the items could vary from frame to frame. This program allows the user to work with either situation.

Input Screen



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 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 this box can be left blank or a "0" (zero) can 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 this box can be left blank or a "0" (zero) can be entered.

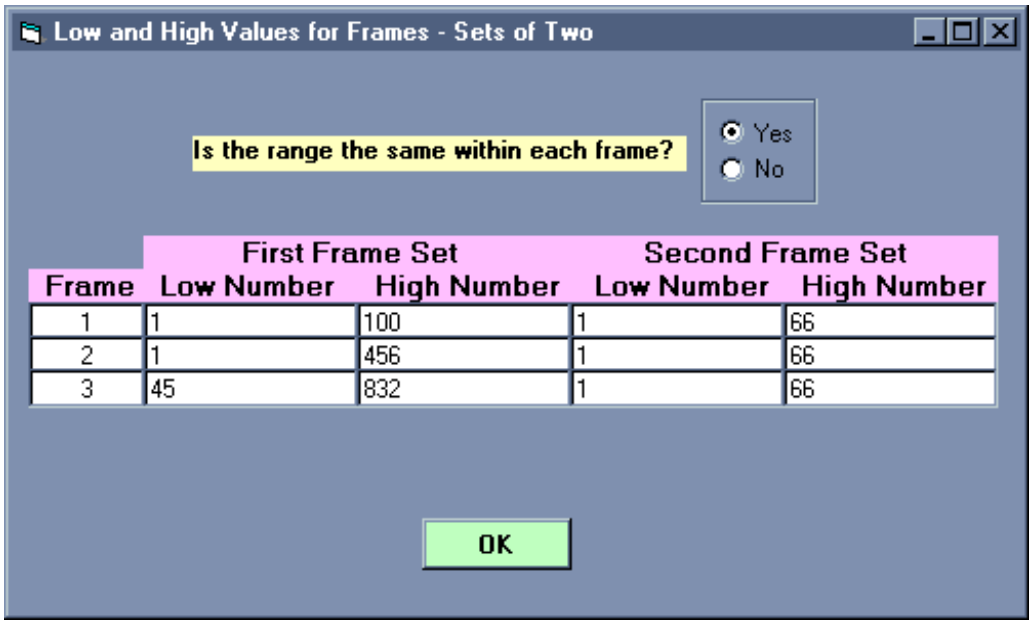
Enter the number of frames in your universe (Max = 51)

A frame is a grouping of contiguous sample items within the universe. In the illustration to follow, the user would have three frames. The first frame would contain pages 1 through 100 in the first volume, the second frame would contain pages 1 through 456 in the second volume, and the third frame would contain pages 45 through 832 in the third volume.

Enter or View low/high values

Based on the frame count entered by the user, the user must enter the low and high boundaries for each frame. When the user clicks on the "Enter or view low/high values" button, the

following form appears. In this illustration, the number of frames is entered as "3" and the "Yes" button is selected in response to the question "Is the range the same within each frame?". In general, there will be one line in this table for each frame in the universe.



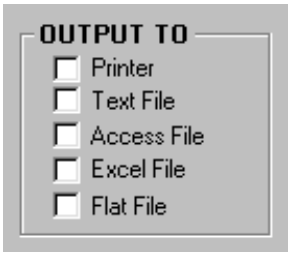
When the "Yes" button is selected in response to the question "Is the range the same within each frame?", the low numbers in the second frame set will be set equal to the low number typed in the first row. For the above illustration, as soon as the user types "1" as the low number for the second frame set in the first frame, the value of "1" will appear in this column for frames 2 and 3 as soon as the user leaves this cell. Similarly, when "66" is typed as the high number for the second frame set in the first frame, the value of "66" will appear in this column for frames 2 and 3 as soon as the user leaves this cell.

If the "No" button is selected in response to the question "Is the range the same within each frame?", the user must supply all values (e.g., 6 for this illustration) for the low and high numbers in the second frame set.

Output Options

The program allows for five output options. The user may select the output 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



Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

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)**. The user will see the standard Windows "Save" screen. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

The output shown below is file C:\ TEMP\OUTDISK.TXT and is the result of generating three random numbers along with two spares using the low/high values in the earlier screen with a fixed range of 66 in each frame.

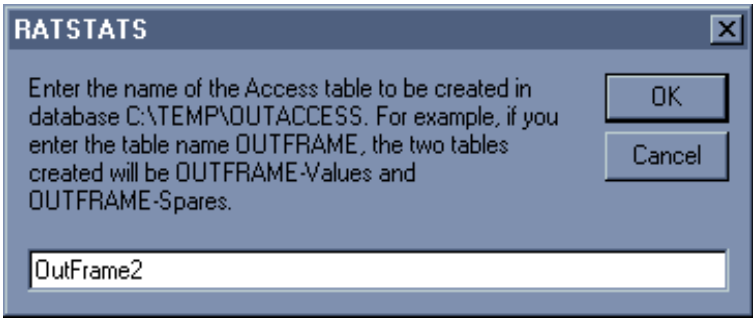
For each random number generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next three pieces of information are a result of converting the random number into the appropriate frame, subframe, and the item within the subframe. In the sequential portion of the output, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/14/2004            Random Number Generator                Time: 9:56
                          AUDIT: User Guide Example

SEED NUMBER: 35809.13                             FRAME SIZE:          88,704


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

     TOTAL RANDOM NUMBERS GENERATED: 5

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:
   POSITIONS  1 THROUGH  6 - ORDER OF SELECTION
   POSITIONS  7 THROUGH 20 - RANDOM NUMBER
   POSITIONS 21 THROUGH 29 - FRAME NUMBER
   POSITIONS 30 THROUGH 42 - SUB-FRAME NUMBER
   POSITIONS 43 THROUGH 55 - ITEM NUMBER WITHIN THE SUB-FRAME
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.


Selection      Random      Frame    Sub-Frame        Item
  Order        Number     Number       Number       Number
     2           7559        2           15           35 ← Three random values start here.
     3          30214        2          358           52
     1          48495        3          223           51
     4          33903        2          414           45 ← The two spare values start here.
     5          40411        3          101           19
```

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

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)**. If the OUTPUT TO **Access File** is selected, the standard Windows "Save" file screen will appear. Fill in the name of the Access database in the **File name** box. The output file will be saved with the ".MDB" extension (e.g., C:\TEMP\OUTACCESS.MDB). By clicking on the **Save** button, the program 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 for this illustration). The table name "OutFrame2" will be used. The program will then create two tables (OutFrame2-Values and OutFrame2-Spares) within database C:\TEMP\OUTACCESS.

The tables OutFrame2-Values and OutFrame2-Spares, shown next, are the result of generating three random numbers along with two spares using the low/high values in the earlier screen with a fixed range of 66 in each frame.

For each random number generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next three pieces of information are a result of converting the random number into the appropriate frame, subframe, and the item within the subframe. The seed number used by the random number generator is in the far-right column. In the sequential portion of the output, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line.

The OutFrame2-Values table contains three random numbers, the seed number, the date, and the time.

OutFrame2-Values : Table

Order	Number	Frame	SubFrame	Item	Seed-Number	Date	Time
2	7559	2	15	35	35809.13	9/7/2005	1:19:00 PM
3	30214	2	358	52			
1	48495	3	223	51			

The OutFrame2-Spares table contains two spare values.

OutFrame2-Spares : Table

Order	Number	Frame	SubFrame	Item
4	33903	2	414	45
5	40411	3	101	19

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

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)**. If the OUTPUT TO **Excel File** is selected, the standard Windows "Save" file screen will appear. Enter the name of the Excel spreadsheet in the **File name** box. The output file will be saved with the ".XLS" extension (e.g., C:\TEMP\OUTEXCEL.XLS).

The following output will be contained in the Excel spreadsheet C:\TEMP\OUTEXCEL.XLS and is the result of generating three random numbers along with two spares using the low/high values in the earlier screen with a fixed range of 66 in each frame. By clicking on the tab labeled "SPARES" the user will obtain the spreadsheet containing the spare values.

For each random number generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next three pieces of information are a result of converting the random number into the appropriate frame, subframe, and the item within the subframe. The output also contains the program execution date and time, the name of the audit/review, the seed number, and the total size of the sampling frame. In the sequential portion of the output, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line.

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

	A	B	C	D	E	F	G
1	Department of Health and Human Services						
2	OIG - Office of Audit Services						
3	Random Number Generator						
4	Date:	10/14/2004	Time:	9:56			
5	Audit:	User Guide Example					
6	Order	Number	Frame	SubFrame	Item	Seed Number	Frame Size
7	2	7559	2	15	35	35809.13	88,704
8	3	30214	2	358	52		
9	1	48495	3	223	51		

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

	A	B	C	D	E
1	Order	Number	Frame	SubFrame	Item
2	4	33903	2	414	45
3	5	40411	3	101	19

NOTE: The user must first exit RAT-STATS in order to view this file.

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

Output to a Flat File

The output immediately following is the text file created by selecting "Flat File" as one of the output options. If the OUTPUT TO **Flat File** is selected, the standard Windows "Save" file screen will appear. 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 flat file is the output file when generating three random numbers along with two spares using the low/high values from the earlier screen with a fixed range of 66 in each frame. 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.

For each random number generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next three pieces of information are a result of converting the random number into the appropriate frame,

subframe, and the item within the subframe. In the sequential portion of the output, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line.

```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/14/2004            Random Number Generator                Time: 9:56
                          AUDIT: User Guide Example

SEED NUMBER: 35809.13                                FRAME SIZE:        88,704


     FILE OF RANDOM NUMBERS: C:\TEMP\OUTFLAT

     TOTAL RANDOM NUMBERS GENERATED: 5

000200000075590002000000001500000000035 ← The three random values start here.
000300000302140002000000035800000000052
000100000484950003000000022300000000051
000400000339030002000000041400000000045 ← The two spare values start here.
000500000404110003000000010100000000019
```

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

Output to a Printer

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

The following printer output is the result of generating three random numbers along with two spares using the low/high values in the earlier screen with a fixed range of 66 in each frame. For each random number generated, five pieces of information are printed. The first is the order of selection. This number indicates the order in which the random number was generated by the program. The next piece of information is the random number generated. The next three pieces of information are a result of converting the random number into the appropriate frame, subframe, and the item within the subframe. In the sequential portion of the output, the lowest random number and the corresponding conversion are printed first. The subsequent numbers are in sequential order with the random number and its equivalent value after conversion on each line.

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/14/2004 RANDOM NUMBER GENERATOR TIME: 9:56
AUDIT: User Guide Example

SEED NUMBER: 35809.13 FRAME SIZE: 88,704

3 RANDOM NUMBERS IN SEQUENTIAL ORDER

ORDER OF SELECTION	RANDOM NUMBER	- - - CONVERTED NUMBER - - - - FRAME	SUB-FRAME	ITEM NUMBER
2 -	7,559	2	15	35
3 -	30,214	2	358	52
1 -	48,495	3	223	51

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

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 10/14/2004 RANDOM NUMBER GENERATOR TIME: 9:56
AUDIT: User Guide Example

SEED NUMBER: 35809.13 FRAME SIZE: 88,704

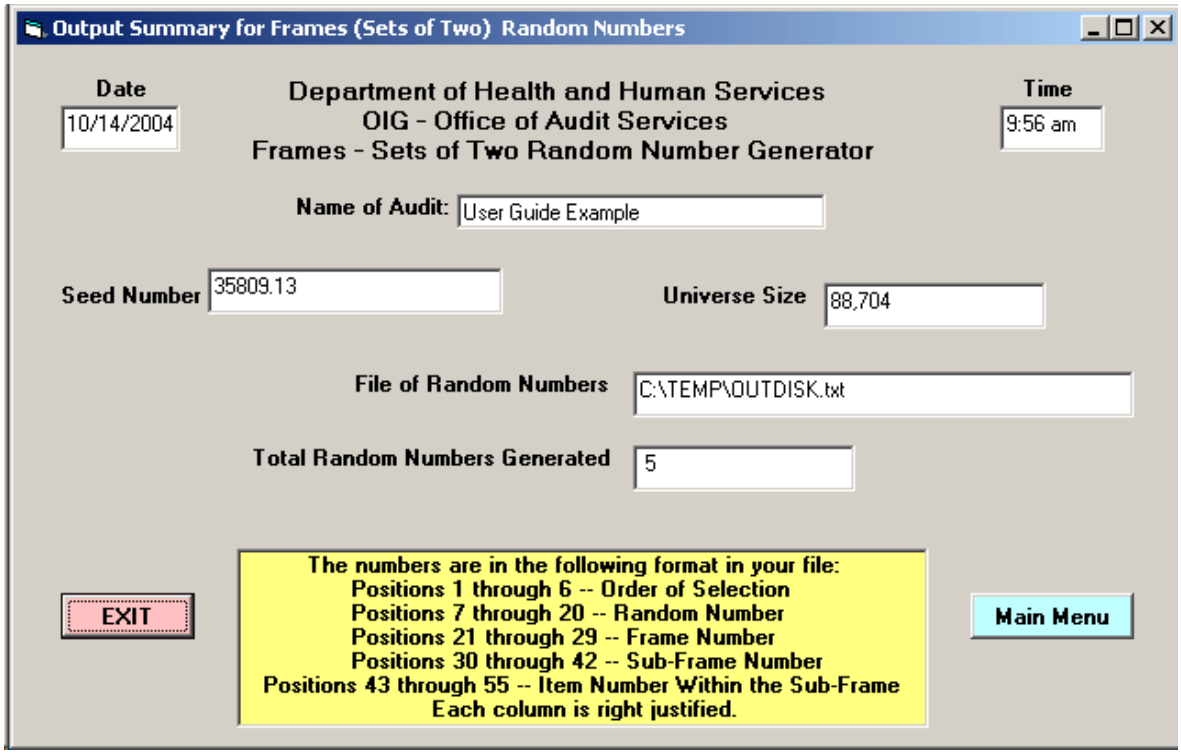
2 RANDOM NUMBERS IN GENERATED ORDER

ORDER OF SELECTION	RANDOM NUMBER	- - - CONVERTED NUMBER - - - - FRAME	SUB-FRAME	ITEM NUMBER
4 -	33,903	2	414	45
5 -	40,411	3	101	19

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

Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



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

RHC SAMPLE SELECTION

Purpose

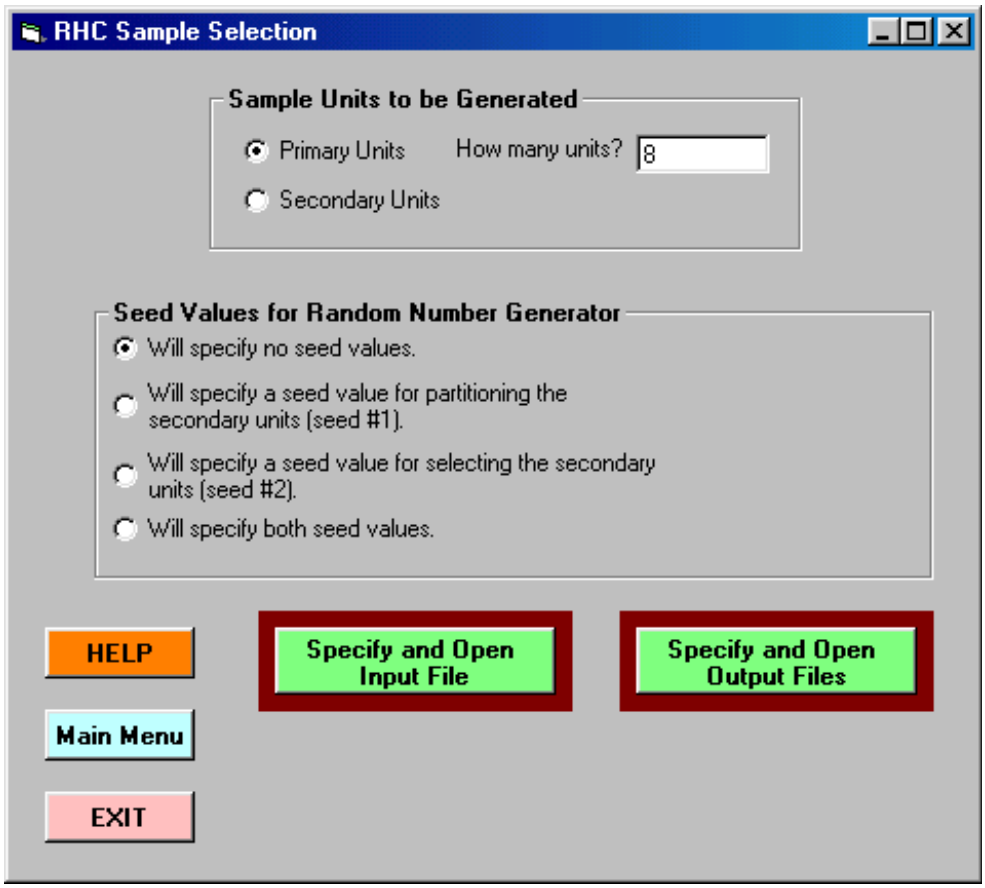
This procedure will select primary units for a two-stage design and both primary and secondary units for a three-stage design using the Rao, Hartley, and Cochran (RHC) methodology. Using this approach, all primary (or secondary if it is a three-stage sample) units are randomly distributed initially into groups. The number of groups is equal to the number of primary (or secondary) units to be sampled. Within each group, one primary (or secondary) unit is sampled based upon probability proportional to size.

The program requires that a data file of all primary (or secondary) units be created. The universe in the data file may not exceed 6,000 units.

The program will identify sub-units for sampling but will not identify the final sample items for review. Another random number program must be used for the final sample selection. For example, the user may want to review Medicare claims at hospitals throughout the United States. The program may first be used to randomly select states. The program may then be used again to randomly select hospitals within the selected states. Another sampling program, such as single stage numbers, could then be used to select the Medicare claims for review. The user does not need to know the size of the universe of hospitals for each state when performing the first phase (i.e., state sampling) of the RHC sample selection. For the second phase, the user must know and create a data file showing information for each hospital in the states sampled. However, the user does not need to know the number of Medicare claims at each hospital in order to perform this phase of the RHC sample selection. Once the sampled hospitals have been identified, the user must determine the quantity of Medicare claims at each hospital selected in order to draw a random sample of claims.

While the program may be used for either the primary or secondary unit selection, for ease of discussion in this section, only the primary will be mentioned.

Input Screen for Selecting Primary/Secondary Units



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

Selecting the Number of Primary/Secondary Units to be Sampled

Enter the number of units to be sampled in the "How many units?" box. The quantity must not be greater than the universe of units. The program will randomly assign the units to groups. The number of groups will be equal to the sample size entered.

Seed Values

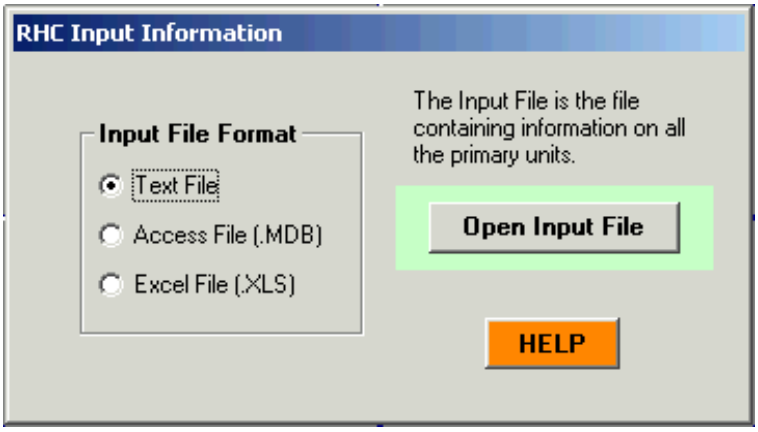
The program allows a seed number to be entered by the user when generating the groups of units. This is called seed #1 in the above screen. This is the first phase of the sample selection. If no number is entered, the program will use an algorithm to generate its own seed number. The seed number algorithm 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 (such as a book of random numbers). The seed number will be displayed as part of the output. If "Will specify a seed value for partitioning the secondary units (seed #1)" is selected, a box will appear on the screen for entering this value. This option allows the user to obtain the same groups of primary units in subsequent computer runs if the same seed value is supplied in this box.

The program allows a seed number to be entered by the user when generating the sample unit from each group of units generated in the first phase. This is called seed #2 in the above screen. If no number is entered, 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 (such as a book of random numbers). The seed number will be displayed as part of the output. If "Will specify a seed value for selecting the secondary units (seed #2)" is selected, a box will appear on the screen for entering this value. This option allows the user to obtain the same sample units from each group of units generated during the first phase in subsequent computer runs if the same seed value is supplied in this box.

Input File Information

To specify information on the input file, click on the button below:

The following form will appear:



Specify the input file format as (1) Text File (any file extension is accepted), (2) Access database (file extension must be .MDB), or (3) Excel spreadsheet (file extension must be .XLS). Next, click on **Open Input File**, select the input file, and click on **Open**.

Format of Input File

The user must create a data file of information on each unit in the universe. If this was an application to generate a sample of secondary units, then information would only be needed on the secondary units of the sampled primary units.

The data for each unit will consist of a description of the unit (e.g., Jefferson County), the quantity of sub-units within the unit (e.g., hospitals in Jefferson County), and a size value for the unit (e.g., number of hospital beds in Jefferson County). The user should try to use a size indicator that most closely relates to the purpose of the sample. For example, while the number of hospital beds in a county may be the best size value for a sample of hospitals, the square footage of hospital space may be the best size indicator in another sample. The size value used affects the variance in the appraisal. However, the sample is not invalidated by using a less-than-best size indicator. The same measure of size must be used throughout the data file. The user, for example, may not use the number of hospital beds for one county and the hospital square footage for another county.

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

NEW YORK 416 18

Explanation:

NEW YORK - This is a description of a unit in the universe. It may be a number or a name related to the unit. There is a maximum limit of 30 characters for the description. Commas, hyphens, and other special characters may be used.

416 - This is the quantity of subunits within a given unit. If the user wanted to sample payroll records at branch offices, this value could represent the number of payroll records at a particular branch office. If the quantity is not known, the user may enter a **1** as the universe size. Commas may be used in entering the numbers.

18 - This is the size value associated with a particular unit. For example, if the user was sampling payroll records at branch offices, this value may represent the number of employees at a particular office. Commas and decimal points may be used in entering the value.

Input from a Text File

If the input file is stored in a text file and the user selects "Text File" in the **Input File Format** list, the standard Windows "Open" file screen will appear. Click on the file name (INTEXT.TXT for this illustration) and click on **Open** (or simply double-click on the file name).

```
COUNTY ONE              1     2990
COUNTY TWO              1      576
COUNTY THREE            1     2086
COUNTY FOUR             1      127
COUNTY FIVE             1      682
COUNTY SIX              1      307
COUNTY SEVEN            1     1226
COUNTY EIGHT            1      301
COUNTY NINE             1    21342
COUNTY TEN              1      166
COUNTY ELEVEN           1      252
COUNTY TWELVE           1      179
COUNTY THIRTEEN         1      493
COUNTY FOURTEEN         1       93
COUNTY FIFTEEN          1      209
```

Input file INTEXT.TXT

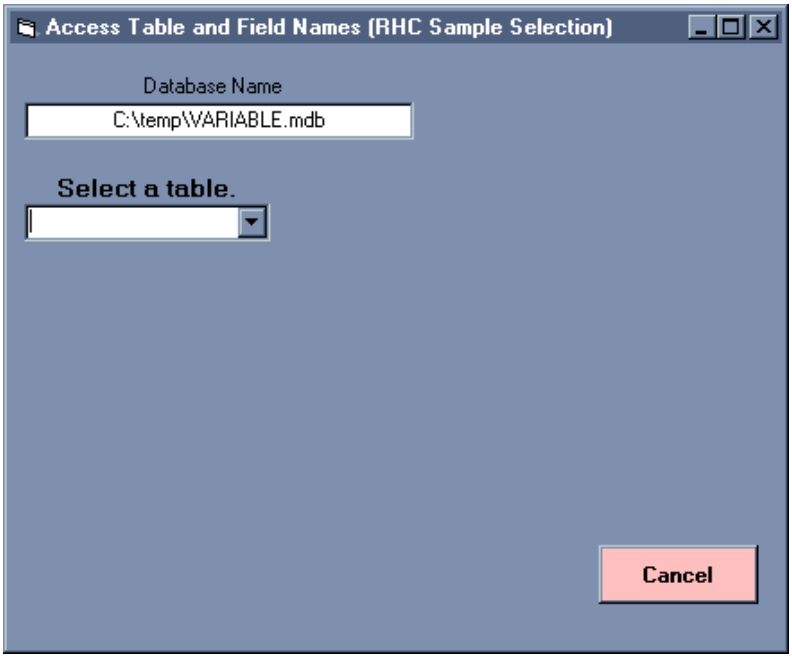
```
COUNTY SIXTEEN          1      163
COUNTY SEVENTEEN        1     1871
COUNTY EIGHTEEN         1     2008
COUNTY NINETEEN         1      253
COUNTY TWENTY           1      300
COUNTY TWENTY-ONE       1     2754
COUNTY TWENTY-TWO       1     2869
COUNTY TWENTY-THREE     1     5202
COUNTY TWENTY-FOUR      1     2557
COUNTY TWENTY-FIVE      1     1770
COUNTY TWENTY-SIX       1      358
COUNTY TWENTY-SEVEN     1      651
COUNTY TWENTY-EIGHT     1      455
COUNTY TWENTY-NINE      1     1772
COUNTY THIRTY           1      250
COUNTY THIRTY-ONE       1      135
COUNTY THIRTY-TWO       1      662
COUNTY THIRTY-THREE     1      344
COUNTY THIRTY-FOUR      1      175
COUNTY THIRTY-FIVE      1      927
COUNTY THIRTY-SIX       1      350
COUNTY THIRTY-SEVEN     1      600
COUNTY THIRTY-EIGHT     1      260
COUNTY THIRTY-NINE      1      502
COUNTY FORTY            1      182
```

Input From an Access Database

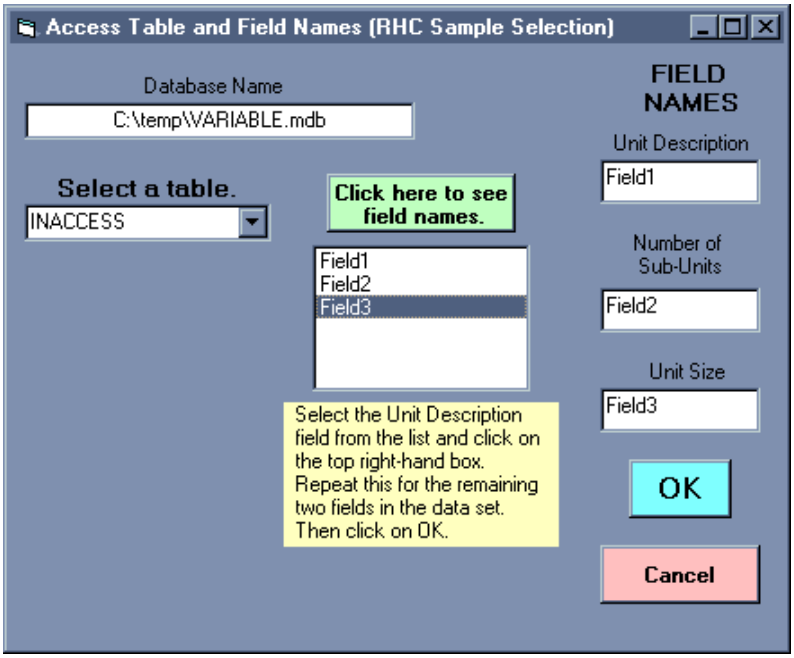
If the input file is stored in a table within an Access database and the user selects "Access File (.MDB)" in the **Input File Format** list, the standard Windows "Open" file screen will appear, listing all Access database (.MDB) files. Click on the database file name (C:\TEMP\VARIABLE.MDB for this illustration) and click on **Open** (or simply double-click on the database file name VARIABLE). Next, 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**.

INACCESS : Table

	Field1	Field2	Field3
	COUNTY ONE	1	2990
	COUNTY TWO	1	576
	COUNTY THREE	1	2086
	COUNTY FOUR	1	127



After selecting a table from the drop-down list and clicking on **Click here to see field names**, the following form will appear:



To select the field names, click on the field name for the field containing the **Unit Description** (Field1 in this illustration) and click on the top right-hand box. The field name will appear in

this box. Repeat this for the field name containing the **Number of Sub-Units** and click on the middle right-hand box to specify this field name (Field2 in this illustration). Repeat using the field name containing the **Unit Size** and specify this field (Field3 in this illustration) by clicking on the bottom right-hand box. When the field names have been specified, click on **OK**. The program will return to the original screen for this module.

Input From an Excel Spreadsheet

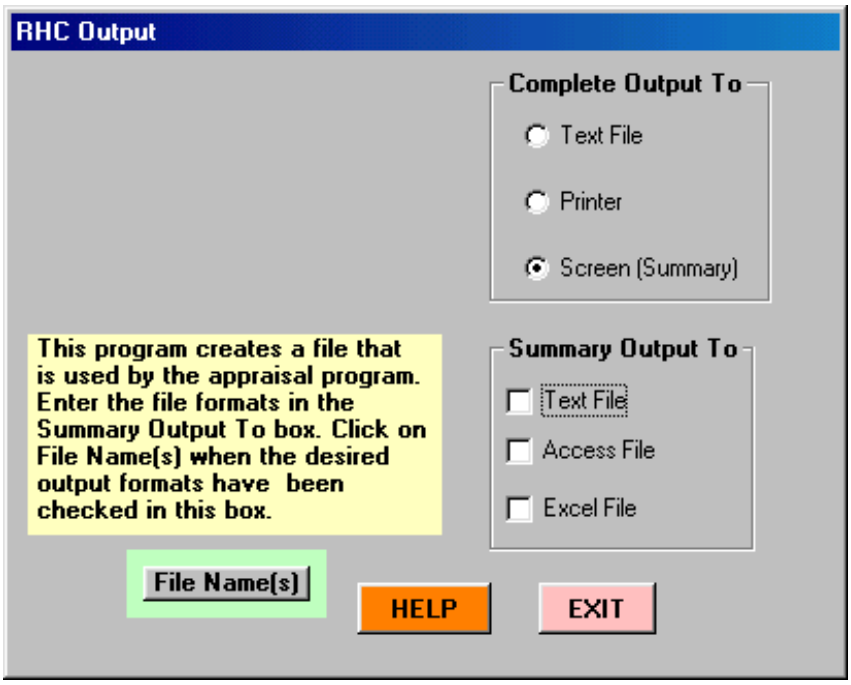
If the input file is stored in an Excel spreadsheet and the user selects "Excel File (.XLS)" in the **Input File Format** list, the standard Windows "Open" file screen will appear, listing all Excel (.XLS) files. Click on the Excel file name (C:\TEMP\INEXCEL.XLS for this illustration) and click on **Open** (or simply double-click on INEXCEL.XLS). To return to the original screen for this module, click on **OK**.

	A	B	C
1	COUNTY ONE	1	2990
2	COUNTY TWO	1	576
3	COUNTY THREE	1	2086
4	COUNTY FOUR	1	127
5	COUNTY FIVE	1	682
6	COUNTY SIX	1	307
7	COUNTY SEVEN	1	1226
8	COUNTY EIGHT	1	301
9	COUNTY NINE	1	21342
10	COUNTY TEN	1	166

Output Information

To specify program output information, click on the button below:

The following form will appear:



Format for Complete Output

This output contains information on the groups of primary/secondary units that were formed and the primary/secondary unit selected from each group. This information can be output to a text file, a printer, or screen. If either "Text File" or "Printer" is selected, the user will be provided very brief summary information in the output screen.

If complete information is desired, the user should select either "Text File" or "Printer." If "Text File" is selected, the standard Windows "Save" file screen will appear. Enter the **File Name** (C:\TEMP\OUTRHC.TXT for this illustration) and click on **Save**. This file will be saved with a ".TXT" extension. If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the appropriate printer and click on **OK**. The program always concludes with a brief summary on the screen.

Format for Summary Output

This file is one of the files needed by the RHC appraisal program which is run after the sample results have been obtained. This file can be in any one or more of the following formats: (1) a Text File, (2) a table within an Access database, or (3) an Excel spreadsheet. When all the desired output formats have been selected in the **Summary Output To** box, click on the button labeled **File Name(s)**.

Text Summary File

If output to **Text File** is selected, the standard Windows "Save" file screen will appear. Type in the **File Name** (C:\TEMP\OUTSUMMARY.TXT for this illustration) and click on **Save**. This file will be saved with a ".TXT" extension. After clicking on **Save**, the program returns to the original screen for this module. To continue processing the data, click on **CONTINUE**.

Access Summary File

If the output to **Access File** is selected, the standard Windows "Save" screen will appear, listing all Access database (.MDB) files. If the database already exists, click on the database name (C:\TEMP\VARIABLE.MDB for this illustration) and click on **Save** (or simply double click on database VARIABLE). If the database does not exist, type in the database name in the **File Name** box and click on **Save**. After clicking on **Save**, the program returns to the original screen for this module. To continue processing the data, click on **CONTINUE**. Next, the user will see the following message asking for the name of the output table within the previously specified Access database. For this illustration, the Access table name is OutSummary.

Excel Summary File

If the output to **Excel File** is selected, the standard Windows "Save" file screen will appear, listing all Excel (.XLS) files. If the spreadsheet already exists and the user wants to replace it, click on the spreadsheet name (C:\TEMP\OUTSUMMARY.XLS for this illustration) and then click on **Save** (or simply double-click on the file OUTSUMMARY.XLS). If the spreadsheet does not exist, enter the spreadsheet name in the **File Name** box and click on **Save**. After clicking on **Save**, the program returns to the original screen for this module. To continue processing the data, click on **CONTINUE**.

Program Output

The **Summary Output to** file contains information needed by the RHC appraisal program which is run after the sample results have been obtained. This file can be in any one or more of the following formats: (1) a Text File, (2) a table within an Access database, or (3) an Excel spreadsheet. The following information is contained in the file:

PRIMARY UNIT ID	This is the description entered by the user for the unit that was selected as a sample item.
SECONDARY UNIVERSE	The number of secondary units entered in the data file for the unit being sampled.
PRIMARY UNIT SIZE	The size factor entered in the data file for the unit being sampled.
GROUP SIZE	The sum of the size factors for the units placed in a particular group. The individual and group size will be used by the RHC appraisal program.
UNITS IN GROUP	The number of units that were randomly placed into the group.

If the user had selected a secondary unit sample, PRIMARY would be replaced by SECONDARY and SECONDARY would be replaced by 3RD STAGE in all the headings shown above.

Important Notice

Versions of Windows RAT-STATS prior to RAT-STATS 2005 used executable DOS modules when generating random numbers. Consequently, the random numbers generated agreed with those obtained using DOS versions of RAT-STATS. RAT-STATS 2005 does not utilize these DOS modules but instead uses the full 32-bit technology of Visual Basic. As a result, random numbers generated using RAT-STATS 2005 and later (which includes RAT-STATS 2007) will not agree with those obtained using any of the previous versions.

Complete Output to Text File or Printer

The output created by the "Text File" option and the "Printer" option is identical and shown below for the sample illustration. This is file C:\TEMP\OUTSUMMARY.TXT specified earlier. Refer to the preceding **Program Output** section for the column descriptions in this output.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/14/2004      GENERATION OF PRIMARY UNIT SAMPLE          Time: 10:35
                    NAME OF INPUT FILE: C:\TEMP\INTEXT.TXT

                          GROUPS OF PRIMARY UNITS

                    ********* GROUP 1 *********
                                          PRIMARY UNIT    SECONDARY
          PRIMARY UNIT IDENTIFICATION         SIZE         UNIVERSE
          ==============================  =============  =============
          COUNTY THIRTY-NINE                        502              1
          COUNTY TWENTY-EIGHT                       455              1
          COUNTY TWO                                576              1
          COUNTY TWENTY-NINE                      1,772              1
          COUNTY SIX                                307              1

               GROUP TOTALS:  5                   3,612              5

                    ********* GROUP 2 *********
                                          PRIMARY UNIT    SECONDARY
          PRIMARY UNIT IDENTIFICATION         SIZE         UNIVERSE
          ==============================  =============  =============
          COUNTY THIRTY-SIX                         350              1
          COUNTY FORTY                              182              1
          COUNTY ONE                              2,990              1
          COUNTY THIRTY-TWO                         662              1
          COUNTY SEVEN                            1,226              1

               GROUP TOTALS:  5                   5,410              5

                    ********* GROUP 3 *********
                                          PRIMARY UNIT    SECONDARY
          PRIMARY UNIT IDENTIFICATION         SIZE         UNIVERSE
          ==============================  =============  =============
          COUNTY FOUR                               127              1
          COUNTY TWENTY-TWO                       2,869              1
          COUNTY FIVE                               682              1
          COUNTY THIRTY                             250              1
          COUNTY THIRTY-ONE                         135              1

               GROUP TOTALS:  5                   4,063              5

                    ********* GROUP 4 *********
                                          PRIMARY UNIT    SECONDARY
          PRIMARY UNIT IDENTIFICATION         SIZE         UNIVERSE
          ==============================  =============  =============
          COUNTY SIXTEEN                            163              1
```

```
COUNTY TWENTY-FOUR                      2,557              1
COUNTY NINETEEN                           253              1
COUNTY TWENTY-THREE                     5,202              1
COUNTY THIRTY-THREE                       344              1

     GROUP TOTALS:  5                   8,519              5

               ********* GROUP 5 *********
                                PRIMARY UNIT    SECONDARY
PRIMARY UNIT IDENTIFICATION          SIZE         UNIVERSE
==============================  =============  =============
COUNTY FOURTEEN                            93              1
COUNTY THIRTY-EIGHT                       260              1
COUNTY TWENTY-SEVEN                       651              1
COUNTY FIFTEEN                            209              1
COUNTY THIRTY-FOUR                        175              1

     GROUP TOTALS:  5                   1,388              5

               ********* GROUP 6 *********
                                PRIMARY UNIT    SECONDARY
PRIMARY UNIT IDENTIFICATION          SIZE         UNIVERSE
==============================  =============  =============
COUNTY EIGHTEEN                         2,008              1
COUNTY TWENTY-SIX                         358              1
COUNTY TEN                                166              1
COUNTY SEVENTEEN                        1,871              1
COUNTY THIRTY-SEVEN                       600              1

     GROUP TOTALS:  5                   5,003              5

               ********* GROUP 7 *********
                                PRIMARY UNIT    SECONDARY
PRIMARY UNIT IDENTIFICATION          SIZE         UNIVERSE
==============================  =============  =============
COUNTY ELEVEN                             252              1
COUNTY TWELVE                             179              1
COUNTY EIGHT                              301              1
COUNTY THREE                            2,086              1
COUNTY NINE                            21,342              1

     GROUP TOTALS:  5                  24,160              5

               ********* GROUP 8 *********
                                PRIMARY UNIT    SECONDARY
PRIMARY UNIT IDENTIFICATION          SIZE         UNIVERSE
==============================  =============  =============
COUNTY TWENTY                             300              1
COUNTY TWENTY-ONE                       2,754              1
COUNTY THIRTY-FIVE                        927              1
COUNTY TWENTY-FIVE                      1,770              1
COUNTY THIRTEEN                           493              1

     GROUP TOTALS:  5                   6,244              5
```

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/14/2004      GENERATION OF PRIMARY UNIT SAMPLE          Time: 10:35
                 NAME OF OUTPUT FILE: C:\TEMP\OutSummary.txt

 FIRST SEED NUMBER:     100.00     SECOND SEED NUMBER:     200.00

     NUMBER OF PRIMARY UNITS IN THE POPULATION:    40
     NUMBER OF PRIMARY UNITS SAMPLED:               8

                                                                      UNITS
                              SECONDARY       PRIMARY                   IN
PRIMARY UNIT ID                UNIVERSE      UNIT SIZE      GROUP SIZE  GROUP
=========================  =============  =============  =============  =====
COUNTY TWENTY-NINE                     1          1,772          3,612      5
COUNTY ONE                             1          2,990          5,410      5
COUNTY TWENTY-TWO                      1          2,869          4,063      5
COUNTY TWENTY-THREE                    1          5,202          8,519      5
COUNTY TWENTY-SEVEN                    1            651          1,388      5
COUNTY SEVENTEEN                       1          1,871          5,003      5
COUNTY NINE                            1         21,342         24,160      5
COUNTY TWENTY-ONE                      1          2,754          6,244      5
```

The last eight lines in the output are the same information contained in the file specified in the **Summary Output To** list. This is one of the files required in the RHC appraisal program.

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

Summary Output to Text File

The file C:\TEMP\OUTSUMMARY.TXT follows. This is identical to the last eight lines of the complete output file.

```
COUNTY TWENTY-NINE          1          1772          3612          5
COUNTY ONE                  1          2990          5410          5
COUNTY TWENTY-TWO           1          2869          4063          5
COUNTY TWENTY-THREE         1          5202          8519          5
COUNTY TWENTY-SEVEN         1           651          1388          5
COUNTY SEVENTEEN            1          1871          5003          5
COUNTY NINE                 1         21342         24160          5
COUNTY TWENTY-ONE           1          2754          6244          5
```

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

Summary Output to Access Database

The table OutSummary is shown below and contains the same information contained in the last eight lines of the complete output file.

OutSummary : Table

Primary-Unit-ID	Secondary-Univ	Primary-Unit-Si	Group-Size	Units-In-Group
COUNTY TWENTY-NINE	1	1772	3612	5
COUNTY ONE	1	2990	5410	5
COUNTY TWENTY-TWO	1	2869	4063	5
COUNTY TWENTY-THREE	1	5202	8519	5
COUNTY TWENTY-SEVEN	1	651	1388	5
COUNTY SEVENTEEN	1	1871	5003	5
COUNTY NINE	1	21342	24160	5
COUNTY TWENTY-ONE	1	2754	6244	5

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

Summary Output to Excel Spreadsheet

The Excel spreadsheet named C:\TEMP\OUTSUMMARY.XLS is shown below and contains the same information contained in the last eight lines of the complete output file.

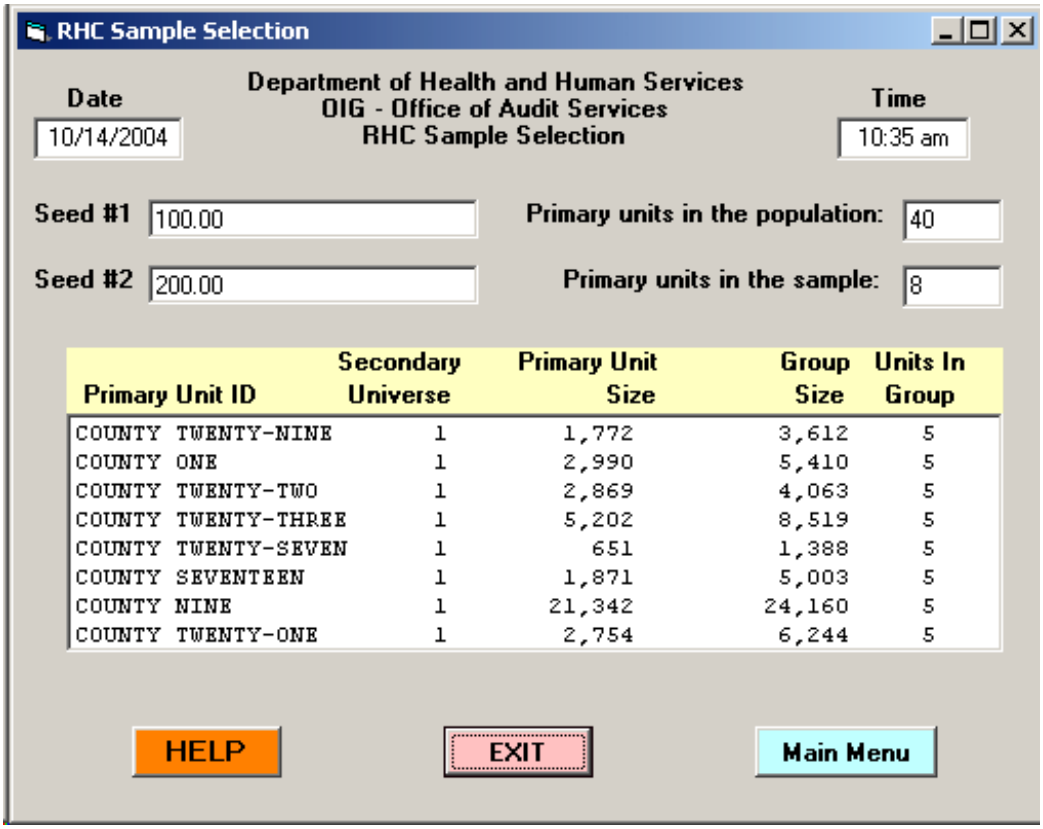
	A	B	C	D	E
1	**Primary Unit ID**	**Secondary Universe**	**Primary Unit Size**	**Group Size**	**Units in Group**
2	COUNTY TWENTY-NINE	1	1772	3612	5
3	COUNTY ONE	1	2990	5410	5
4	COUNTY TWENTY-TWO	1	2869	4063	5
5	COUNTY TWENTY-THREE	1	5202	8519	5
6	COUNTY TWENTY-SEVEN	1	651	1388	5
7	COUNTY SEVENTEEN	1	1871	5003	5
8	COUNTY NINE	1	21342	24160	5
9	COUNTY TWENTY-ONE	1	2754	6244	5

NOTE: The user must first exit RAT-STATS in order to view this file.

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

Summary Output to Screen

The program always concludes with a summary on the screen. The following screen is the summary for this illustration:



Primary Unit ID	Secondary Universe	Primary Unit Size	Group Size	Units In Group
COUNTY TWENTY-NINE	1	1,772	3,612	5
COUNTY ONE	1	2,990	5,410	5
COUNTY TWENTY-TWO	1	2,869	4,063	5
COUNTY TWENTY-THREE	1	5,202	8,519	5
COUNTY TWENTY-SEVEN	1	651	1,388	5
COUNTY SEVENTEEN	1	1,871	5,003	5
COUNTY NINE	1	21,342	24,160	5
COUNTY TWENTY-ONE	1	2,754	6,244	5

This screen contains the two seed values and the number of sampled primary/secondary units. The contents of the large box consist of the same information contained in the files specified in the **Program Output** list.

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

Attribute Appraisals

OVERVIEW

UNRESTRICTED

STRATIFIED

TWO-STAGE UNRESTRICTED

THREE-STAGE UNRESTRICTED

RHC TWO STAGE

RHC THREE STAGE

STRATIFIED CLUSTER

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 an approving signature is on a document or as complicated as to whether 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 appraisal methodologies used 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 included 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 timecards from a pay period to determine the number of timecards that show employees arriving late for work. An unrestricted random sample of timecards 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 timecard review mentioned above, the reviewer may want to place each timecard 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.

Two-Stage Unrestricted

The cost of performing a review may cause the reviewer to use a sampling method that could help reduce the costs. For example, the timecards mentioned above could be located throughout the United States. The reviewer could randomly select locations throughout the organization and then, at the selected locations, select a sample of timecards. This sample design could reduce travel costs for the reviewer.

Three-Stage Unrestricted

This methodology is similar to the "Two-Stage Unrestricted" with the addition of another level of sampling. Using the timecard example, the reviewer may decide to select regions of the country, then sample locations within the selected regions. Then at the selected locations, the user would select timecards.

RHC Two Stage

In certain situations a user may want to draw a multistage sample with the probability greater for selecting "larger" units in the universe. For example, a user may want to take an inventory of items at various warehouses. However, the user may want the larger warehouses to have a greater chance for selection. The RHC statistical methodology allows the user to weight the primary units (e.g., by using square footage at each warehouse) and thereby increase the chance for larger units to be sampled. It provides a method of sample selection that allows sampling without replacement while resembling the use of probability proportional to size sampling.

When the primary units are selected, the size of each primary unit is considered rather than obtaining a simple random sample of primary units.

The primary units are selected using the RHC Sample Selection program. A random sample is then obtained for each selected primary unit and the number of secondary units having the attribute of interest (e.g., in error) is recorded.

RHC Three Stage

This is similar to the two-stage RHC procedure in that the primary units are selected by considering the size of each unit. In addition, the size of the secondary units is considered when selecting them from the sampled primary units.

The primary and secondary units are selected using the RHC Sample Selection program. A random sample of third-stage units is then obtained for each sampled primary/secondary unit combination and the number of third-stage units having the attribute of interest (e.g., in error) is recorded.

Stratified Cluster

The selection of a sample item may sometimes be extremely costly in terms of time and resources. However, once the item has been selected it can be reviewed rather quickly. For example, using the timecards again, once the reviewer has arrived at the sampled location, it may take only a short amount of additional time to review all timecards as compared to a sample of the cards. The reviewer may decide to group the locations by number of employees, with large locations having more than 100 employees. From each group (stratum) the reviewer would sample locations. At each selected location, the reviewer would analyze all the timecards.

Stratified Multistage

This methodology is similar to the "Stratified Cluster" with the exception that not all items in the subuniverse are reviewed. In the example discussed above, the assumption was made that the reviewer had sufficient time and resources to analyze all of the timecards at the selected locations. This may not be possible. Therefore, this methodology is used when a sample is still needed at each selected location.

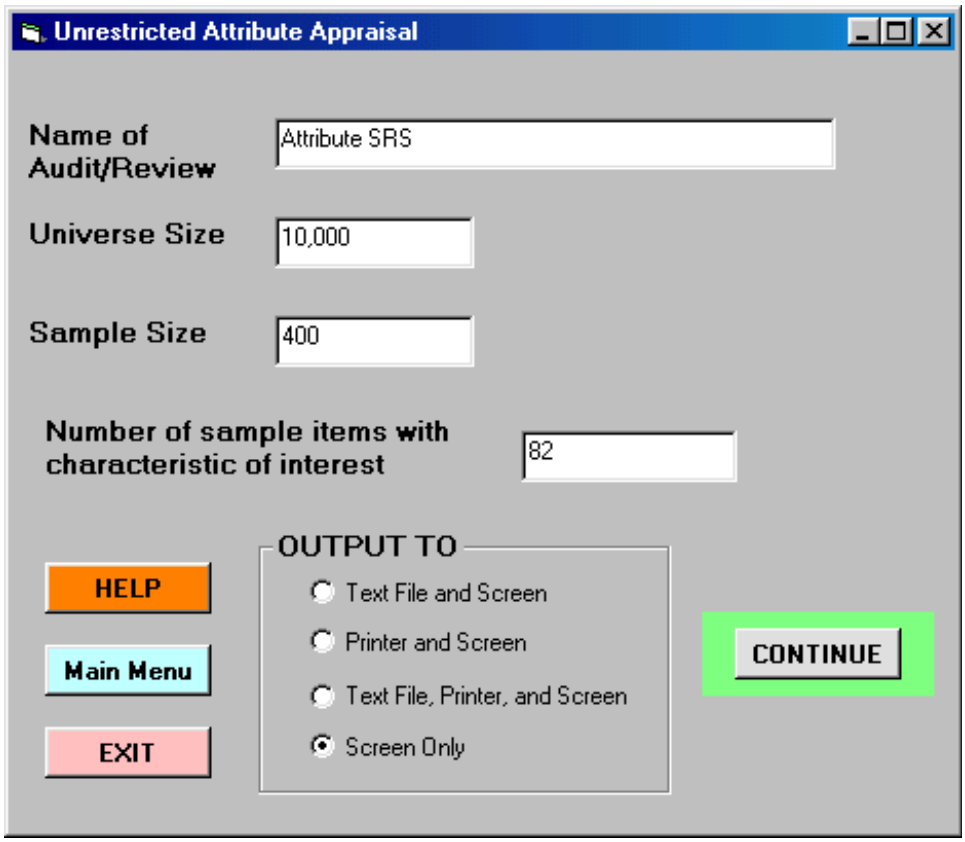
These appraisal programs provide the correct statistical results only when the proper sample design has been executed. If an unrestricted sample of timecards were drawn from throughout the organization, only the unrestricted attribute appraisal program would generate the proper results. Therefore, the reviewer must be sure at the time the sample approach is developed that the appropriate appraisal methodology will be used.

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 to 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.

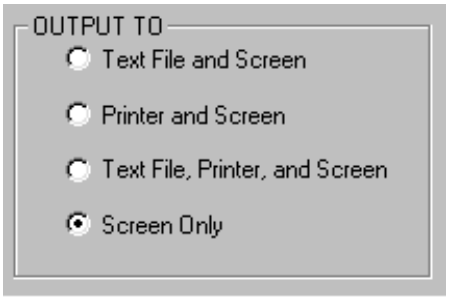
Number of sample items with characteristic of interest

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 or 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 the 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.

Output Options

The program allows for three types of output. The user may select the output to be sent to a text file, printer, or screen. The user selects the appropriate output by clicking the corresponding button.



The output options are a text file and screen; a printer and screen: a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. By selecting the appropriate printer and clicking on **OK**, the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

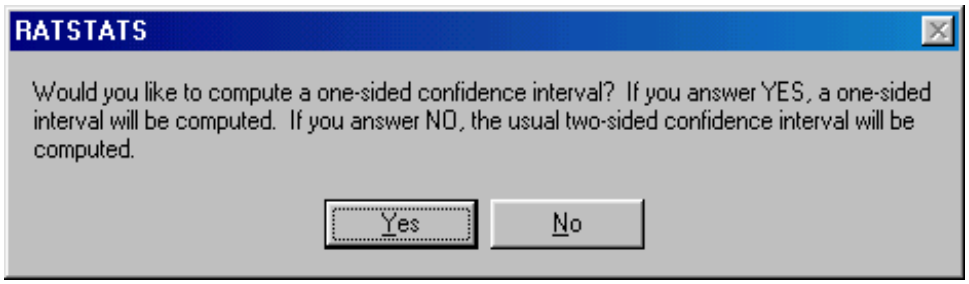
Program Output

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 precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following information will also 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.
STANDARD ERROR	An estimate of the standard deviation of the point estimate for the proportion of sample items and the universe total having the characteristic of interest. This is a measure of the sample precision.

CONFIDENCE LEVELS	The confidence (80%, 90%, or 95%) that the actual proportion (or total number in the universe) will fall within the corresponding confidence interval.
LOWER LIMIT	The lower boundary of the confidence interval. The limit is shown as both a number and 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 percentage of the universe. The confidence levels are 80%, 90%, and 95%.

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 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: 10/22/2004         Single Stage Attribute Appraisal          Time: 10:48
                          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                                                      1.980%

                                     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%
```

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

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."

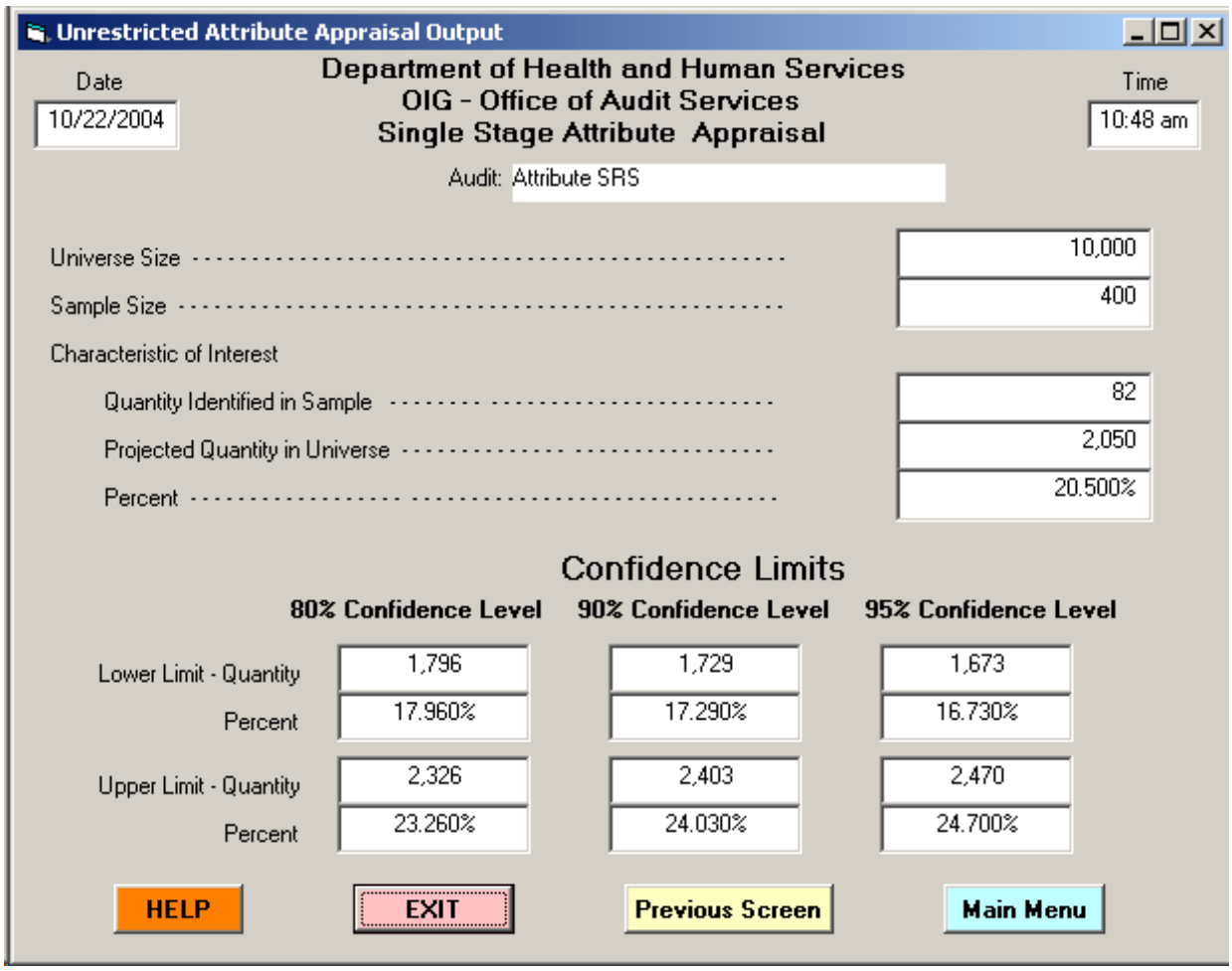
```
                    Department of Health and Human Services
                         OIG - Office of Audit Services
Date: 10/22/2004       Single Stage Attribute Appraisal            Time: 10:52
                         AUDIT/REVIEW: Attribute SRS


         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."

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

Output to Screen

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



Unrestricted Attribute Appraisal Output

Date: 10/22/2004

Department of Health and Human Services
OIG - Office of Audit Services
Single Stage Attribute Appraisal

Time: 10:48 am

Audit: Attribute SRS

Universe Size	10,000
Sample Size	400
Characteristic of Interest	
Quantity Identified in Sample	82
Projected Quantity in Universe	2,050
Percent	20.500%

Confidence Limits

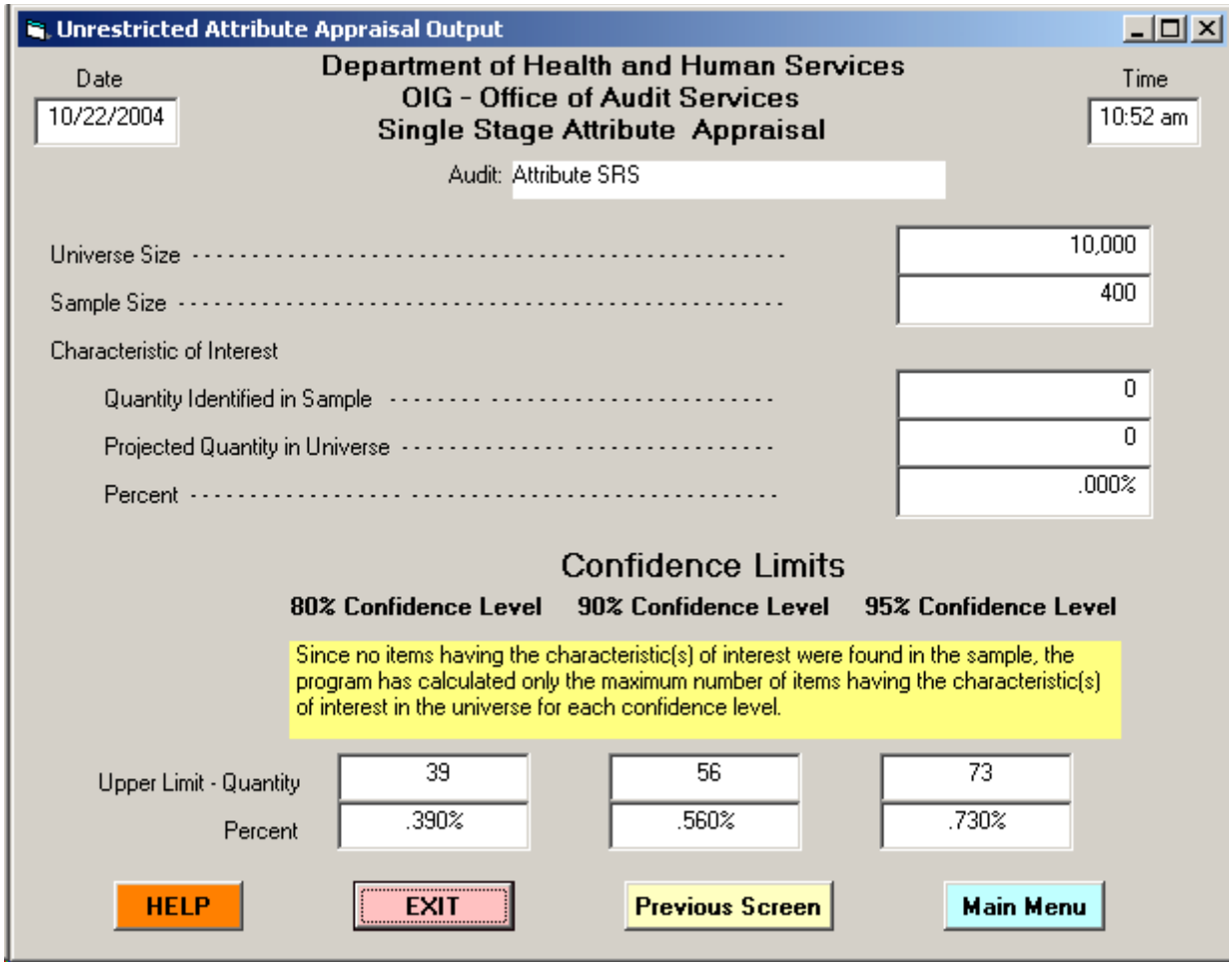
	80% Confidence Level	90% Confidence Level	95% Confidence Level
Lower Limit - Quantity	1,796	1,729	1,673
Percent	17.960%	17.290%	16.730%
Upper Limit - Quantity	2,326	2,403	2,470
Percent	23.260%	24.030%	24.700%

HELP EXIT Previous Screen Main Menu

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

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."

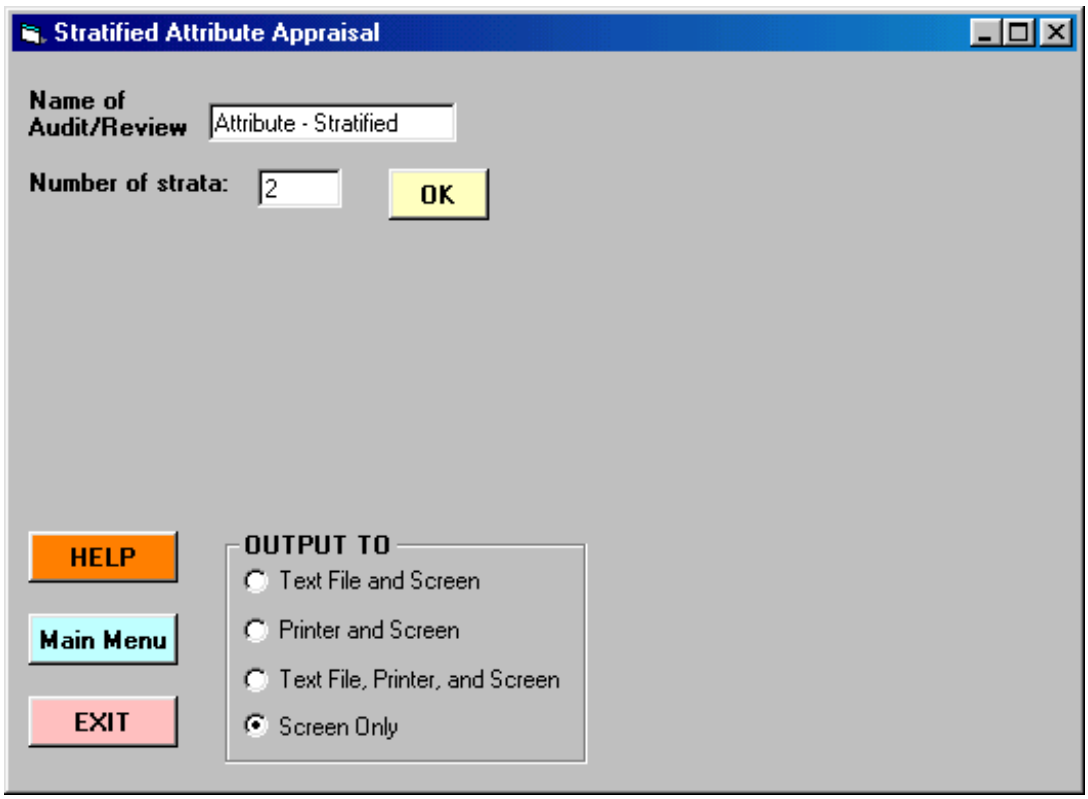
> **NOTE**: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

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



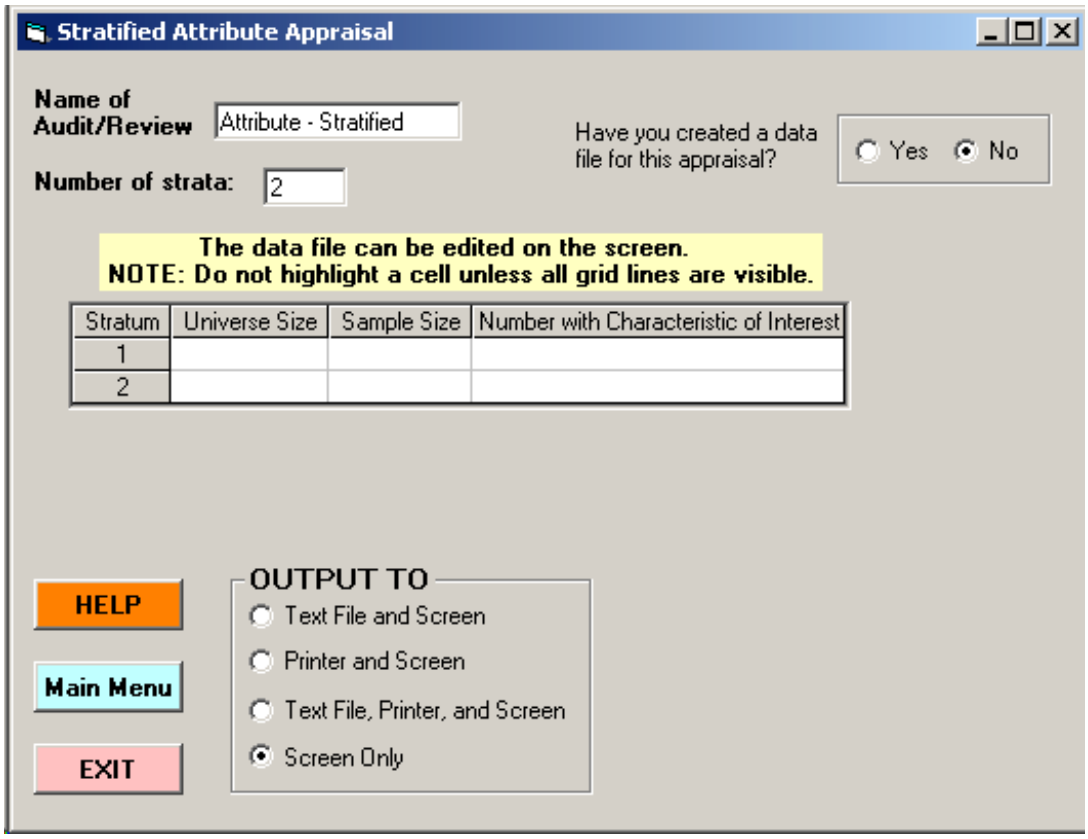
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

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.

Number of strata

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



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

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, then click on **Open**. 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.

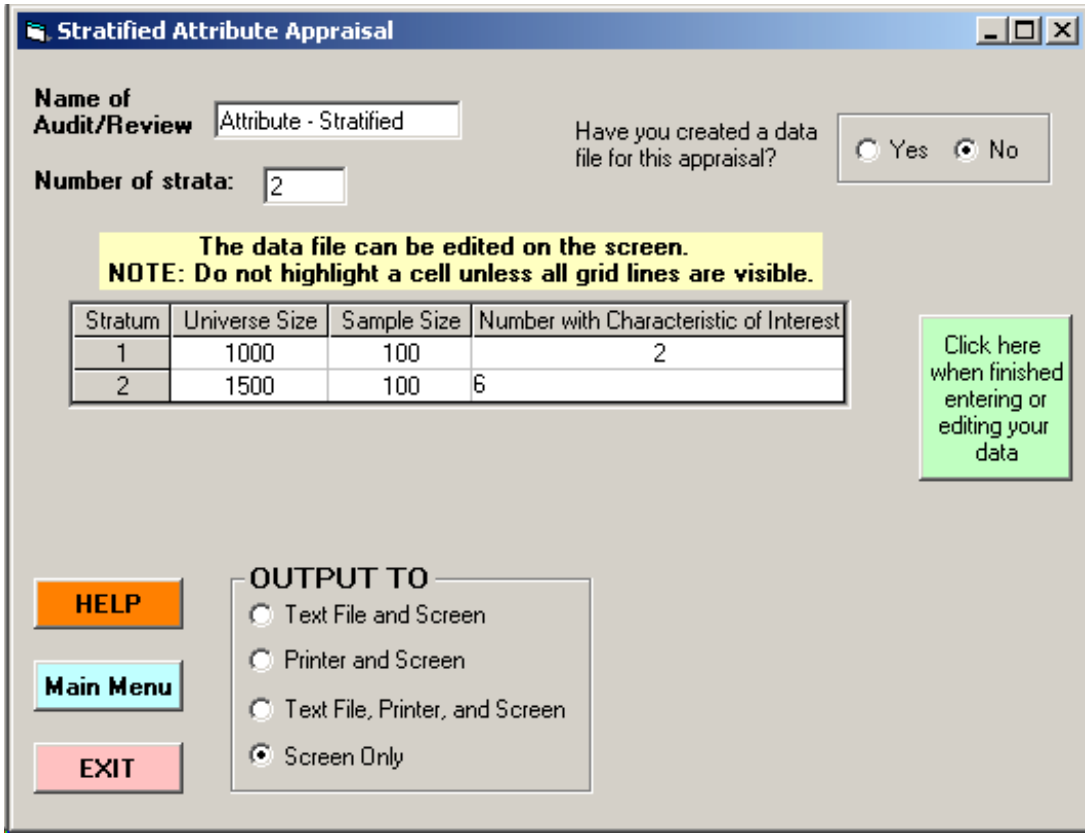
2 - This is the quantity of items that met the criteria established for the sample review.

The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (e.g., 10,000) since the program will assume that the 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**: 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.



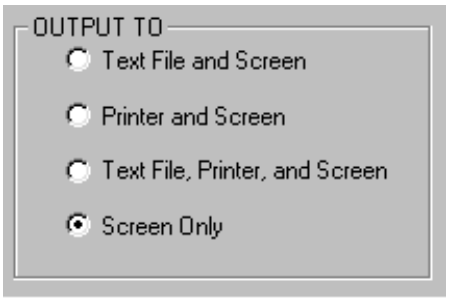
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

When all the values within the grid have been entered, click on **Click here when finished entering or editing your data**, the **Save Input Data** button will appear.

To save this data set, click on **Save Input 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**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the appropriate printer and click on **OK** and the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The program will output the data supplied by the user (universe size, number of items with characteristic of interest, and sample size) for each stratum. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following information will also 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.

STANDARD ERROR	A measurement of the standard deviation of the sample proportion with respect to all possible proportions for this universe and sample size.
PRECISION	The confidence interval half-width expressed as a percentage. Precision is calculated for each stratum and the universe as a whole.
CONFIDENCE LEVEL (CL)	The confidence (80%, 90%, or 95%) that the actual proportion (or total number in the universe) will fall within the corresponding confidence interval.
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 or Printer

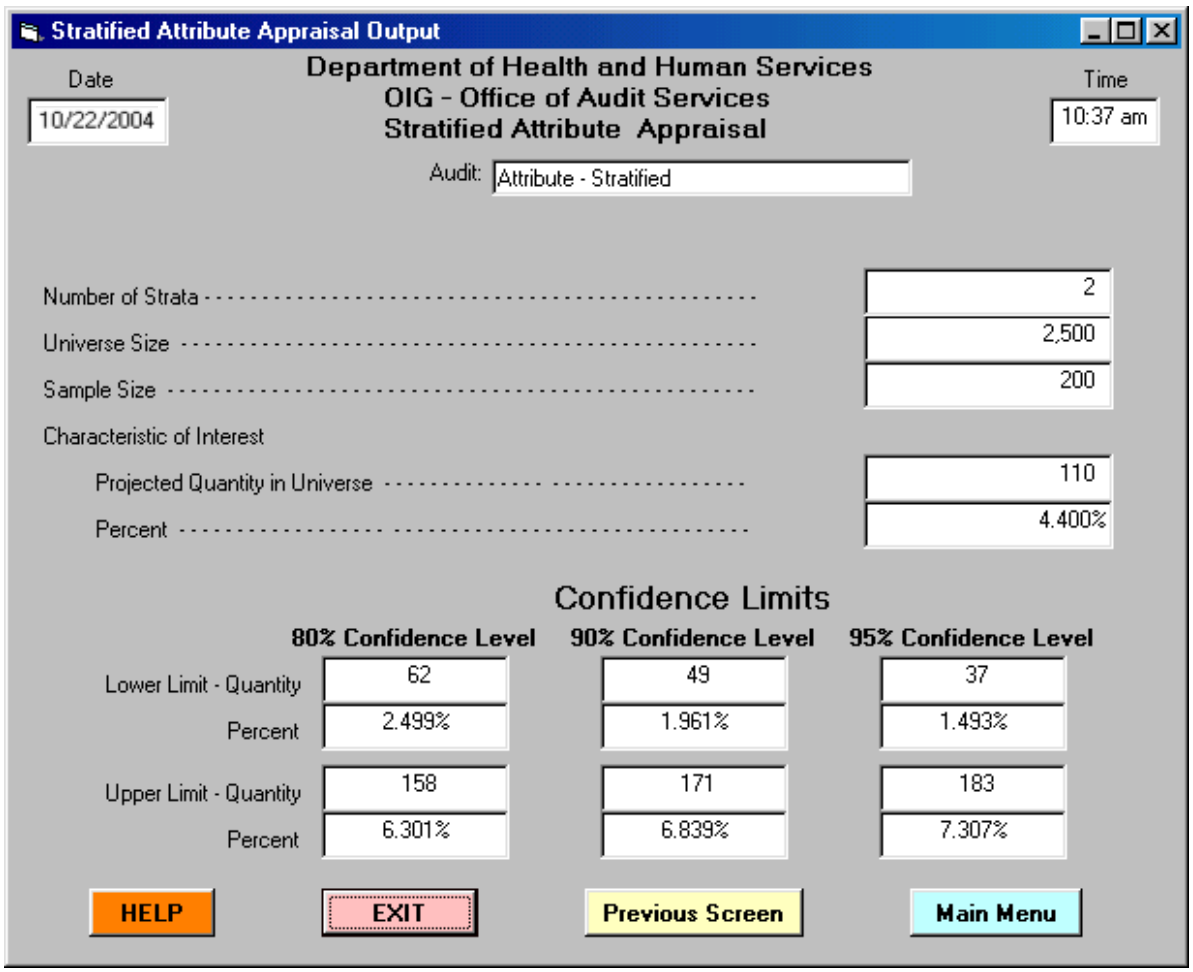
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.TXT, shown next. The printer output is identical but does not include the output file name if the output was not also saved to a text file. Both the text file and printer output will include the name of the input file if the data were input in this manner.

```
              DEPARTMENT OF HEALTH & HUMAN SERVICES
                 OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004   STRATIFIED ATTRIBUTE APPRAISAL     Time: 10:37
              AUDIT/REVIEW: Attribute - Stratified
                OUTPUT FILE: C:\TEMP\OUTDISK.TXT

                                                     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


                            PRECISION     PRECISION    PRECISION
STRATUM                     AT 80% CL     AT 90% CL    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%
```

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

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. This summary will also include the name of the input file if the data were entered in this manner.



Stratified Attribute Appraisal Output

Date: 10/22/2004

Department of Health and Human Services
OIG - Office of Audit Services
Stratified Attribute Appraisal

Time: 10:37 am

Audit: Attribute - Stratified

Number of Strata	2
Universe Size	2,500
Sample Size	200
Characteristic of Interest	
Projected Quantity in Universe	110
Percent	4.400%

Confidence Limits

	80% Confidence Level	90% Confidence Level	95% Confidence Level
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%

HELP EXIT Previous Screen Main Menu

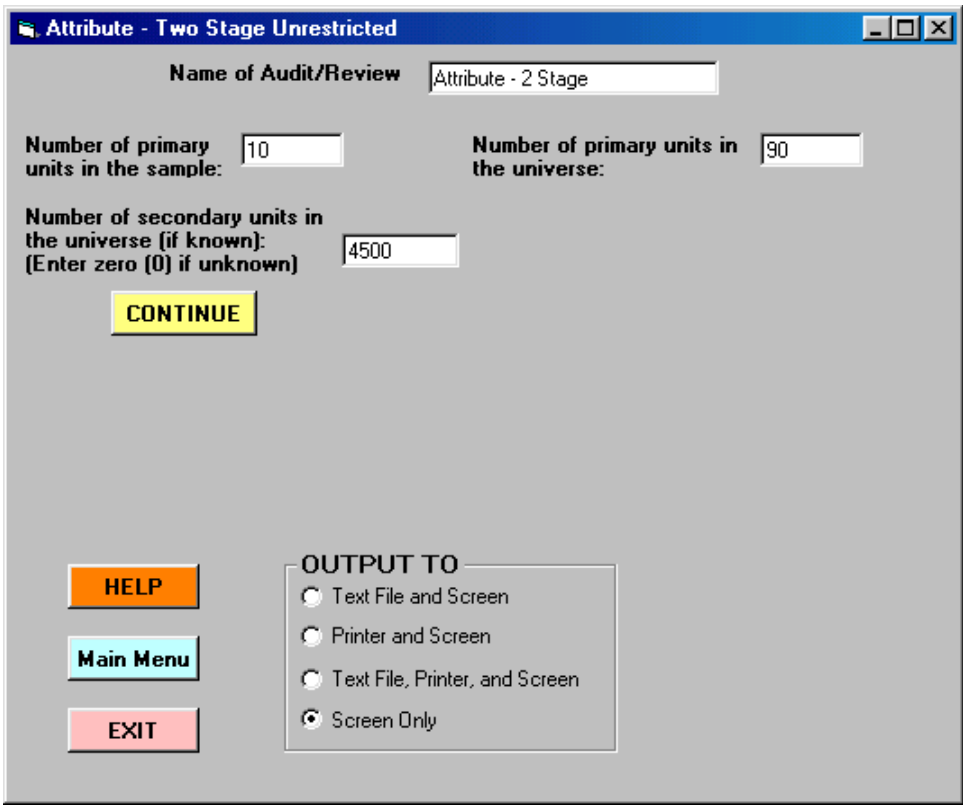
NOTE: Example is for illustrative purposes only. The sample size may not conform to the organization's minimum sample size standards.

TWO-STAGE UNRESTRICTED

Purpose

This program provides a two-stage 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. The program will accept a maximum of 150 primary sampling units for an appraisal.

Input Screen



NOTE: Example is for illustrative purposes only. The sample size may not conform to 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.

Number of primary units in the universe:

This is the number of primary units in the universe from which the primary items were sampled. For example, the primary universe could be counties within a state. From a universe of 58 counties, the user might sample 12 counties. The response to this query would be 58.

Number of primary units in the sample:

Enter the number of sampled primary units in this box. For example, from a universe of 58 counties, the user sampled 12 counties. The response to this query would be 12.

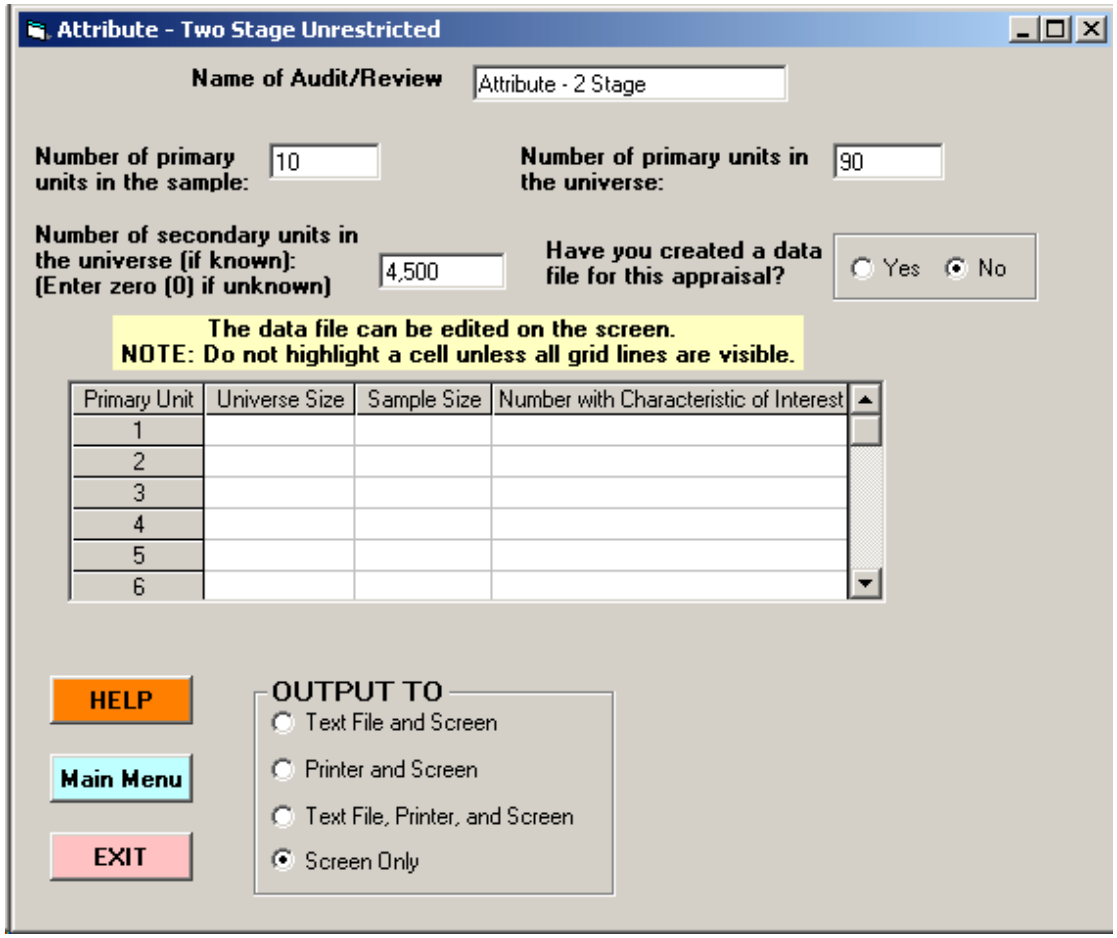
Number of secondary units in the universe (if known):

This is the total number of all transactions or events (secondary units) in the universe. In the above example, secondary units could be claims in each county. The response to this query would be the claim count for all 58 counties. This total number of secondary items may be difficult to determine. Therefore, the user may enter a zero if the quantity is unknown.

If a zero is entered, a different set of formulas will be used that utilize a ratio-type estimator. This estimator is statistically biased, but the bias is negligible if the number of sampled primary units is large.

If the total number of secondary units is entered, the formulas use an unbiased estimator.

When all values have been specified, 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 primary unit, 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. Click on the file name (DATA2STG.TXT for this illustration) and click on **Open** (or simply double-click on file DATA2STG.TXT).

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:

50 10 4

Explanation:

50 - This is the number of items in the primary unit from which a sample was drawn.

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

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

The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (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 value that contains nonnumeric data.

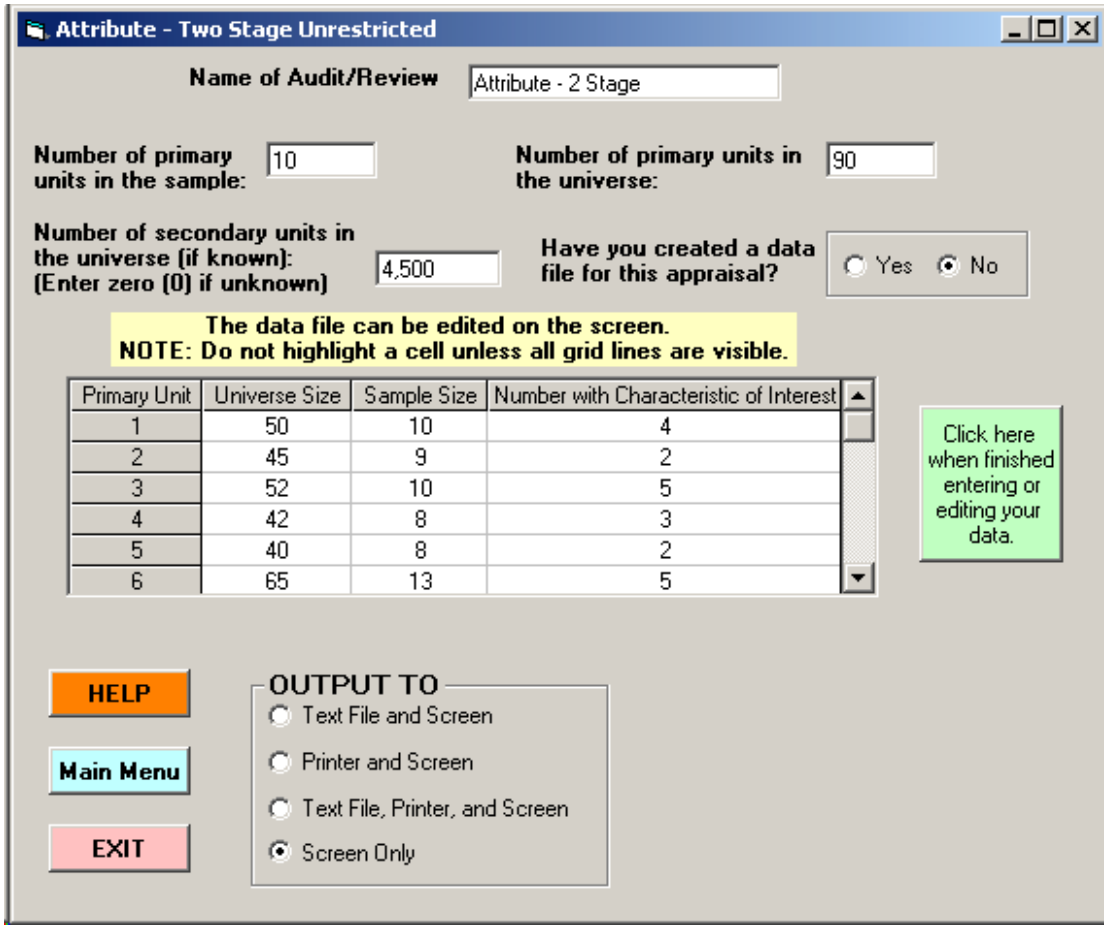
The text file C:\TEMP\DATA2STG.TXT used in this illustration is shown below. Ten primary units are sampled from a total of 90. The number of secondary units in the universe is known to be 4,500.

```
50   10    4
45    9    2
52   10    5
42    8    3
40    8    2
65   13    5
48   10    3
58   12    3
66   13    4
56   11    4
```

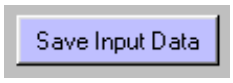
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

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. To illustrate, suppose the first primary unit contains 50 universe values, 10 of which are sampled. Also, the number of sample items meeting the criteria (e.g., in error) is 4. These values should be entered as shown on the screen. **NOTE**: This grid structure will 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.



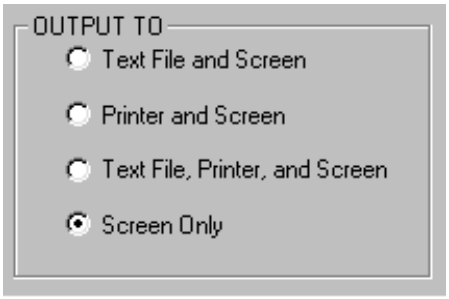
When all the values within the grid have been entered, click on **Click here when finished entering or editing your data** and the **Save Input Data** button will appear.



To save this data set, click on **Save Input 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**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the appropriate printer and click on **OK** and the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The program will print the data supplied by the user (universe size, sample size, and number of sample items with characteristic of interest) for each primary unit. Overall totals are shown for the number of primary units in the universe and the number of secondary items in the universe (if

known). The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following information will also be displayed for each primary unit:

RATIO	The percentage of sample items in each primary unit with the characteristic of interest.
PROJECTED	If the total number of secondary items in the universe was entered, then the projected number of items meeting the characteristic(s) of interest for each primary unit will be displayed.

The following overall statistics will also be displayed.

OVERALL RATIO	The estimate of the universe proportion, stated as a percentage.
OVERALL PROJECTED	The proportion of sample items with the characteristic of interest multiplied by the number of secondary items in the universe (if known).
STANDARD ERROR	A measurement of the standard deviation of the sample proportion with respect to all possible proportions for this universe and sample size.
CONFIDENCE LEVELS	The confidence (80%, 90%, or 95%) that the actual proportion (or total number in the universe) will fall within the corresponding confidence interval.
LOWER LIMIT	The lower limit of the 80%, 90%, and 95% confidence interval. The limit is shown as both a proportion of the universe and quantity of items if the user entered a non-zero value for the universe of secondary items.
UPPER LIMIT	The upper limit of the 80%, 90%, and 95% confidence interval. The limit is shown as both a proportion of the universe and quantity of items if the user entered a non-zero value for the universe of secondary items.

Output to a Text File or Printer

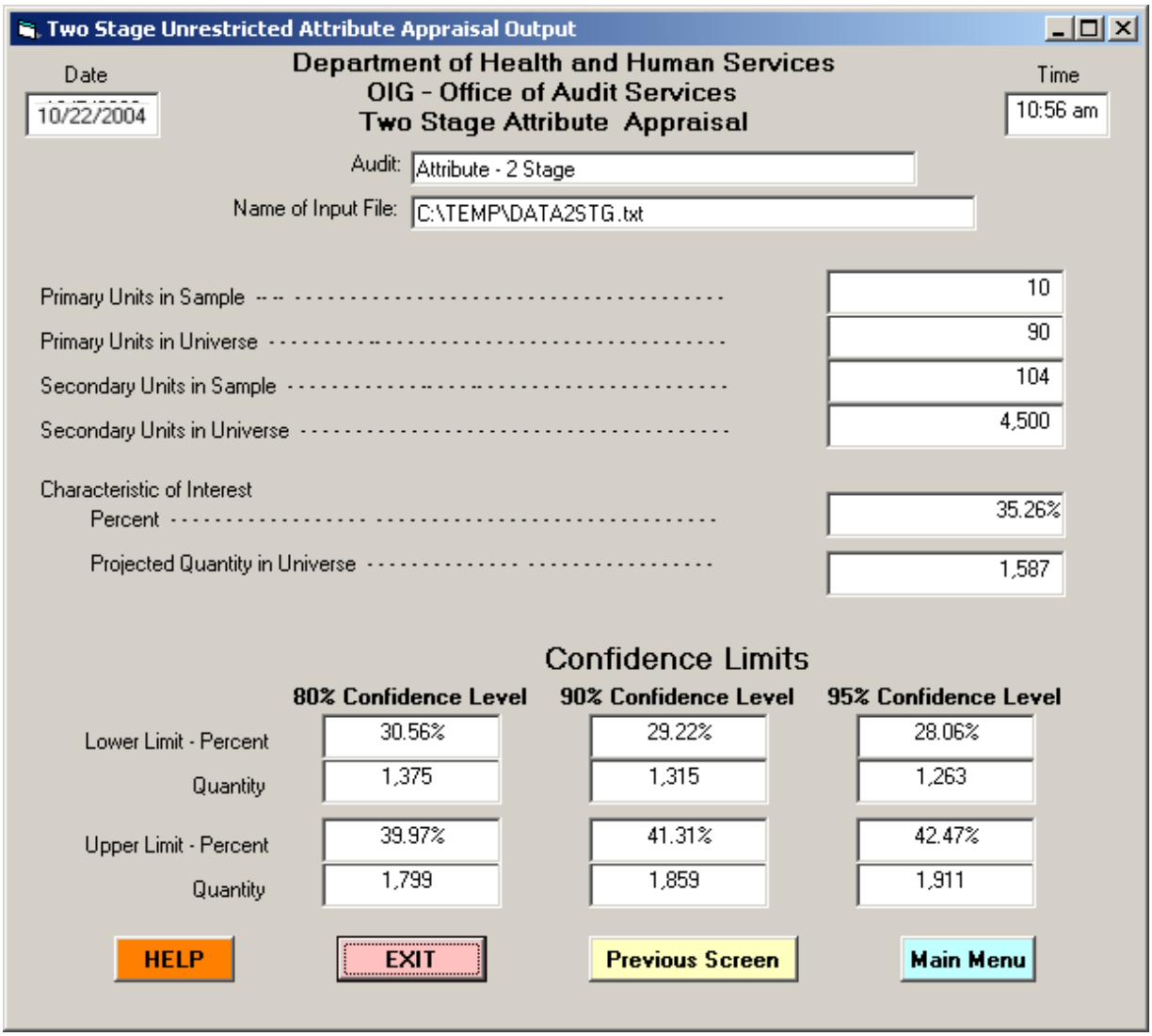
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.TXT, shown next. The printer output is identical but does not include the output file name if the output was not also saved to a text file. Both the text file and printer output will include the name of the input file if the data were input in this manner.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004   TWO STAGE UNRESTRICTED ATTRIBUTE APPRAISAL     Time: 10:56
                      AUDIT/REVIEW: Attribute - 2 Stage
                        DATA FILE: C:\TEMP\DATA2STG.TXT
                       OUTPUT FILE: C:\TEMP\OUTDISK.TXT

PRIMARY                                SAMPLE ITEMS WITH
 UNIT        UNIVERSE    SAMPLE SIZE  CHARACTERISTIC(S)  RATIO      PROJECTED
=======    ============  ===========  =================  =====      =========

   1                50           10              4     40.00%             20
   2                45            9              2     22.22%             10
   3                52           10              5     50.00%             26
   4                42            8              3     37.50%             16
   5                40            8              2     25.00%             10
   6                65           13              5     38.46%             25
   7                48           10              3     30.00%             14
   8                58           12              3     25.00%             15
   9                66           13              4     30.77%             20
  10                56           11              4     36.36%             20

TOTALS             522          104             35

OVERALL TOTALS
  90             4,500                                 35.26%          1,587

STANDARD ERROR                                           3.67%            165

CONFIDENCE LEVEL                80 PERCENT        90 PERCENT        95 PERCENT
LOWER LIMIT FOR PROPORTION          30.56%            29.22%            28.06%
UPPER LIMIT FOR PROPORTION          39.97%            41.31%            42.47%

LOWER LIMIT FOR TOTAL                1,375             1,315             1,263
UPPER LIMIT FOR TOTAL                1,799             1,859             1,911
```

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

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:



Two Stage Unrestricted Attribute Appraisal Output

Date: 10/22/2004

Department of Health and Human Services
OIG - Office of Audit Services
Two Stage Attribute Appraisal

Time: 10:56 am

Audit: Attribute - 2 Stage

Name of Input File: C:\TEMP\DATA2STG.txt

Primary Units in Sample	10
Primary Units in Universe	90
Secondary Units in Sample	104
Secondary Units in Universe	4,500
Characteristic of Interest Percent	35.26%
Projected Quantity in Universe	1,587

Confidence Limits

	80% Confidence Level	90% Confidence Level	95% Confidence Level
Lower Limit - Percent	30.56%	29.22%	28.06%
Quantity	1,375	1,315	1,263
Upper Limit - Percent	39.97%	41.31%	42.47%
Quantity	1,799	1,859	1,911

HELP EXIT Previous Screen Main Menu

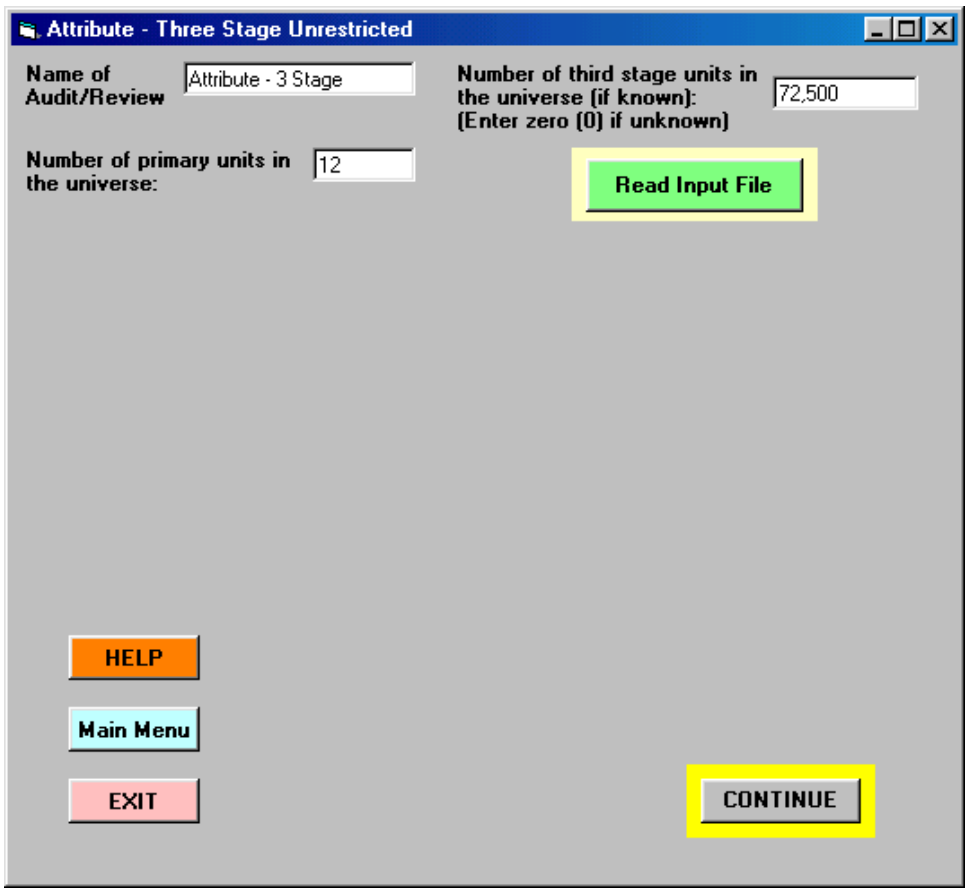
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

THREE-STAGE UNRESTRICTED

Purpose

This program computes the point estimate, standard error, and confidence limits for the universe percentage and universe total number when using a three-stage attribute sampling procedure. Attribute sampling is used to determine how frequently an event or type of transaction occurs in a given universe. The results can be reported as a percentage estimate or an estimate of the total number in the universe having this characteristic of interest. The program will accept a maximum of 50 first-stage units and 150 second-stage units within each first-stage unit (i.e., 50 x 150 = 7,500 second-stage units).

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 primary units in the universe:

This is the number of primary units in the universe from which the primary items were sampled. For example, the primary universe could be regions within the United States. From a universe of 12 regions, one might sample four regions. The response to this query would be 12.

Number of third-stage units in the universe (if known):

If the number of third-stage units in the universe is known, enter this value. If this universe size is known, an unbiased estimate of the universe proportion is obtained. If this universe size is unknown, enter a value of zero, in which case a ratio-type estimator of the universe proportion is used by the program. This estimator is biased, but the bias is negligible if the number of sampled primary units is large.

Read input file

When all values have been specified, click on the **Read Input File** button.

For this program, the user *must* enter the data from a file (i.e., screen input is not allowed). After clicking on **Read Input File**, the standard Windows "Open" file screen will appear. Click on the file name (DATA3ST.TXT for this illustration) and click on **Open** (or simply double-click on file DATA3ST.TXT).

The data file containing the information must 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 storing the data in a spreadsheet software application (e.g., Excel) as a print file. The data file needs to be formatted in the following manner:

REGION 5 90 10
HOSPITAL 1 47 9 3

Explanation:

REGION 5 - is the description of a primary unit. The description may be of any length up to 25 characters and include characters, digits, spaces, and punctuation marks.

90 - is the universe of secondary units within the primary unit.

10 - is the number of secondary units that are being sampled in the primary unit.

HOSPITAL 1 - is a description of a secondary unit that was sampled from the primary unit. The description may be of any length up to 25 characters and include characters, digits, spaces, and punctuation marks.

47 - is the number of third-stage items comprising the universe for the secondary sampled unit.

9 - is the number of third-stage items that were sampled.

3 - is the number of third-stage sampled items that met the characteristic of interest to the user.

The last piece of data (e.g., 10) on each primary unit line is used by the program to determine the number of lines with secondary unit information. The program, therefore, expects that number of lines (e.g., 10) to be lines with secondary unit information. Each line of data must contain information on only one primary unit or one secondary unit. Values within a line can be separated by commas, one or more spaces, or tabs. Commas should not be used within numeric values (e.g., use 2500 rather than 2,500).

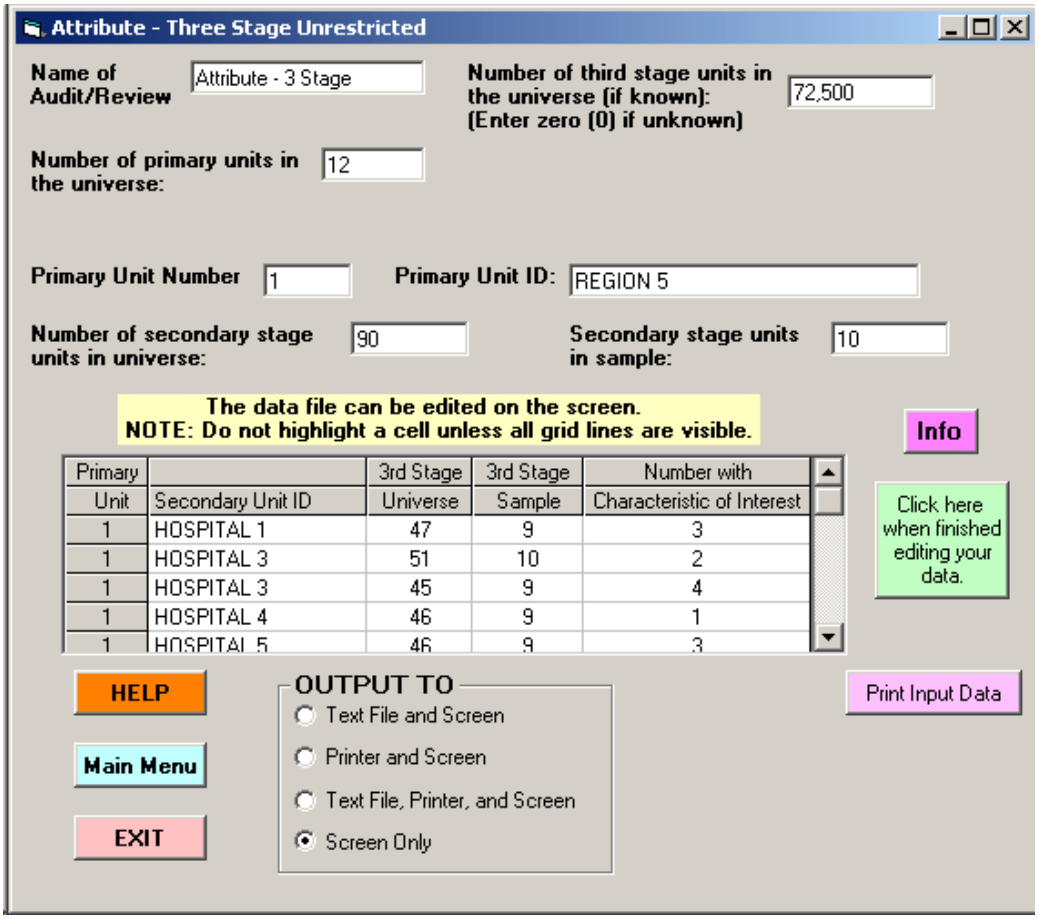
The data file used in this illustration is stored in C:\TEMP\DATA3ST.TXT and is shown below. The primary units are regions within the U.S. and the secondary units are hospitals within each region.

```
REGION 5    90  10
HOSPITAL 1   47   9  3
HOSPITAL 2   51  10  2
HOSPITAL 3   45   9  4
HOSPITAL 4   46   9  1
HOSPITAL 5   46   9  3
HOSPITAL 6   50  10  1
HOSPITAL 7   50  10  4
```

```
HOSPITAL 8   57  11  3
HOSPITAL 9   54  11  4
HOSPITAL 10  64  13  2
REGION 7   110  10
HOSPITAL 1   53  11  2
HOSPITAL 2   59  12  5
HOSPITAL 3   52  10  1
HOSPITAL 4   67  13  3
HOSPITAL 5   59  12  1
HOSPITAL 6   73  15  6
HOSPITAL 7   51  10  3
HOSPITAL 8   75  15  2
HOSPITAL 9   66  13  1
HOSPITAL 10  58  12  4
REGION 8    85  10
HOSPITAL 1   45   9  3
HOSPITAL 2   39   8  2
HOSPITAL 3   43   9  4
HOSPITAL 4   34   7  1
HOSPITAL 5   54  11  2
HOSPITAL 6   54  11  3
HOSPITAL 7   34   7  1
HOSPITAL 8   59  12  1
HOSPITAL 9   49  10  4
HOSPITAL 10  43   9  2
REGION 10  120  10
HOSPITAL 1   59  12  2
HOSPITAL 2   68  14  6
HOSPITAL 3   57  11  3
HOSPITAL 4   72  14  6
HOSPITAL 5   70  14  1
HOSPITAL 6   73  15  2
HOSPITAL 7   83  17  5
HOSPITAL 8   89  18  4
HOSPITAL 9   73  15  3
HOSPITAL 10  77  15  2
```

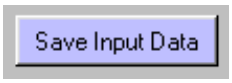
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

After opening the data set, the remaining portion of the input screen will appear.



The primary unit data (e.g., Primary Unit Number, Primary Unit ID, Number of secondary stage units in universe and Secondary stage units in sample) cannot be changed on this screen. For more information, click on the **Info** button. The user may scroll through the secondary unit data or edit any of those cells by clicking on the cell to be changed. **NOTE**: This grid structure will 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.

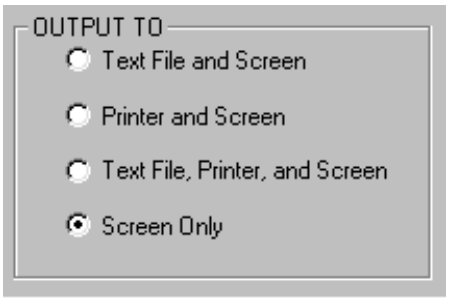
If any values within the grid are edited, click on **Click here when finished editing your data**, the **Save Input Data** button will appear.



To save this data set, click on **Save Input 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**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the appropriate printer and click on **OK**, and the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The first section of the program output contains the name of the data file along with the contents of this file. For each secondary unit in the file, the value of RATIO is printed, where RATIO is the percentage of third-stage sample units within this secondary unit having the characteristic of interest. Totals are shown for each primary unit summarizing (1) the number of third-stage units in the universe, (2) the number of sampled third-stage units, and (3) the number of third-stage units having the characteristic of interest.

The next section of output, entitled OVERALL TOTALS, summarizes the above three values (across all primary units in the sample) along with the number of primary and secondary units in the universe and sample. The universes, with respect to the second and third stages, reflect the populations for stages reviewed and not the universes across all stages. For example if 10 primary units were sampled from a population of 100 primary units, the universe of secondary units displayed in the output represents only the quantity of secondary units in the 10 primary units. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following information is also presented:

OVERALL POINT ESTIMATE OF THE PROPORTION	This is the estimate of the universe proportion having the characteristic of interest.
OVERALL STANDARD ERROR (PROPORTION)	An estimate of the standard deviation of the proportion estimate; a value which determines the width of the corresponding confidence intervals for the parameter.
OVERALL POINT ESTIMATE OF UNIVERSE TOTAL	The estimate of the total number of elements in the universe having the characteristic of interest.
OVERALL STANDARD ERROR (TOTAL)	An estimate of the standard deviation of the universe total estimate; a value which determines the width of the corresponding confidence intervals for the parameter.
CONFIDENCE LEVELS	The confidence (80%, 90%, or 95%) that the actual proportion (or total number in the universe) will fall within the corresponding confidence interval.
LOWER LIMIT FOR PROPORTION	The lower limit of the 80%, 90%, and 95% confidence interval for the universe proportion.
UPPER LIMIT FOR PROPORTION	The upper limit of the 80%, 90%, and 95% confidence interval for the universe proportion.
LOWER LIMIT FOR TOTAL	The lower limit of the 80%, 90%, and 95% confidence interval for the universe total number.
UPPER LIMIT FOR TOTAL	The upper limit of the 80%, 90%, and 95% confidence interval for the universe total number.

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 immediately following. The printer output is identical but does not include the output file name if the output was not also saved to a text file.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004        THREE STAGE ATTRIBUTE APPRAISAL            Time: 11:13
                      AUDIT/REVIEW: Attribute - 3 Stage
                   NAME OF INPUT FILE: C:\TEMP\DATA3ST.txt
                      OUTPUT FILE: C:\TEMP\OUTDISK.TXT
```

FIRST STAGE SECOND STAGE	NEXT STAGE UNIVERSE	SAMPLE SIZE	MEETING CRITERIA	RATIO
REGION 5	90	10		
HOSPITAL 1	47	9	3	33.33%
HOSPITAL 2	51	10	2	20.00%
HOSPITAL 3	45	9	4	44.44%
HOSPITAL 4	46	9	1	11.11%
HOSPITAL 5	46	9	3	33.33%
HOSPITAL 6	50	10	1	10.00%
HOSPITAL 7	50	10	4	40.00%
HOSPITAL 8	57	11	3	27.27%
HOSPITAL 9	54	11	4	36.36%
HOSPITAL 10	64	13	2	15.38%
TOTALS	510	101	27	
REGION 7	110	10		
HOSPITAL 1	53	11	2	18.18%
HOSPITAL 2	59	12	5	41.67%
HOSPITAL 3	52	10	1	10.00%
HOSPITAL 4	67	13	3	23.08%
HOSPITAL 5	59	12	1	8.33%
HOSPITAL 6	73	15	6	40.00%
HOSPITAL 7	51	10	3	30.00%
HOSPITAL 8	75	15	2	13.33%
HOSPITAL 9	66	13	1	7.69%
HOSPITAL 10	58	12	4	33.33%
TOTALS	613	123	28	

REGION 8	85	10		
HOSPITAL 1	45	9	3	33.33%
HOSPITAL 2	39	8	2	25.00%
HOSPITAL 3	43	9	4	44.44%
HOSPITAL 4	34	7	1	14.29%
HOSPITAL 5	54	11	2	18.18%
HOSPITAL 6	54	11	3	27.27%
HOSPITAL 7	34	7	1	14.29%
HOSPITAL 8	59	12	1	8.33%
HOSPITAL 9	49	10	4	40.00%
HOSPITAL 10	43	9	2	22.22%
TOTALS	454	93	23	
REGION 10	120	10		
HOSPITAL 1	59	12	2	16.67%
HOSPITAL 2	68	14	6	42.86%
HOSPITAL 3	57	11	3	27.27%
HOSPITAL 4	72	14	6	42.86%
HOSPITAL 5	70	14	1	7.14%
HOSPITAL 6	73	15	2	13.33%
HOSPITAL 7	83	17	5	29.41%
HOSPITAL 8	89	18	4	22.22%
HOSPITAL 9	73	15	3	20.00%
HOSPITAL 10	77	15	2	13.33%
TOTALS	721	145	34	

TOTAL NUMBER OF THIRD STAGE UNITS IN UNIVERSE 72,500

OVERALL TOTALS	UNIVERSE	SAMPLED
FIRST STAGE	12	4
SECOND STAGE	405{}	40
THIRD STAGE	2,298{}	462
SAMPLED ITEMS MEETING CRITERIA		112

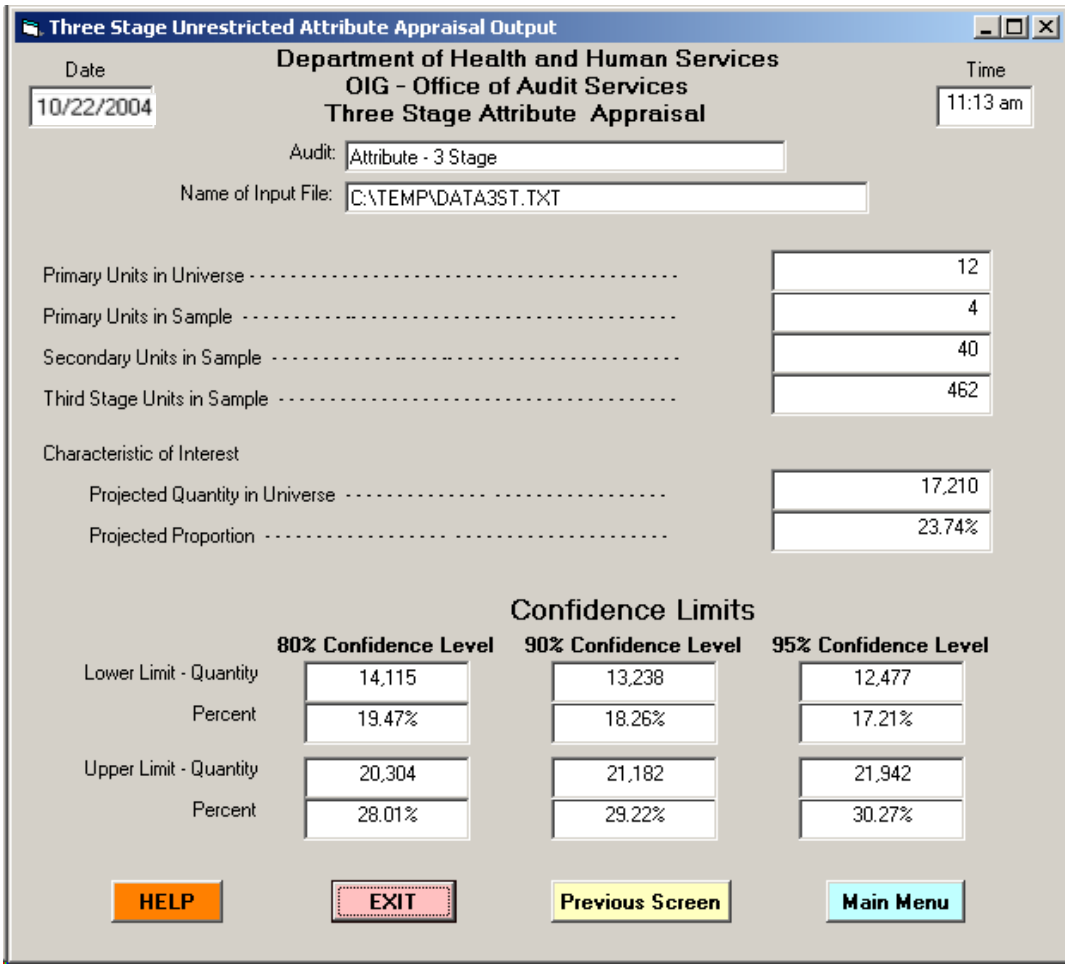
{} UNIVERSE SIZES FOR THE SECOND AND THIRD STAGES REPRESENT THE UNIVERSES FOR THE SAMPLED PRIOR STAGE.

OVERALL POINT ESTIMATE OF THE PROPORTION	23.74%
OVERALL STANDARD ERROR (PROPORTION)	3.33%
OVERALL POINT ESTIMATE OF UNIVERSE TOTAL	17,210
OVERALL STANDARD ERROR (TOTAL)	2,415

CONFIDENCE LEVEL	80 PERCENT	90 PERCENT	95 PERCENT
LOWER LIMIT FOR PROPORTION	19.47%	18.26%	17.21%
UPPER LIMIT FOR PROPORTION	28.01%	29.22%	30.27%
LOWER LIMIT FOR TOTAL	14,115	13,238	12,477
UPPER LIMIT FOR TOTAL	20,304	21,182	21,942

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:



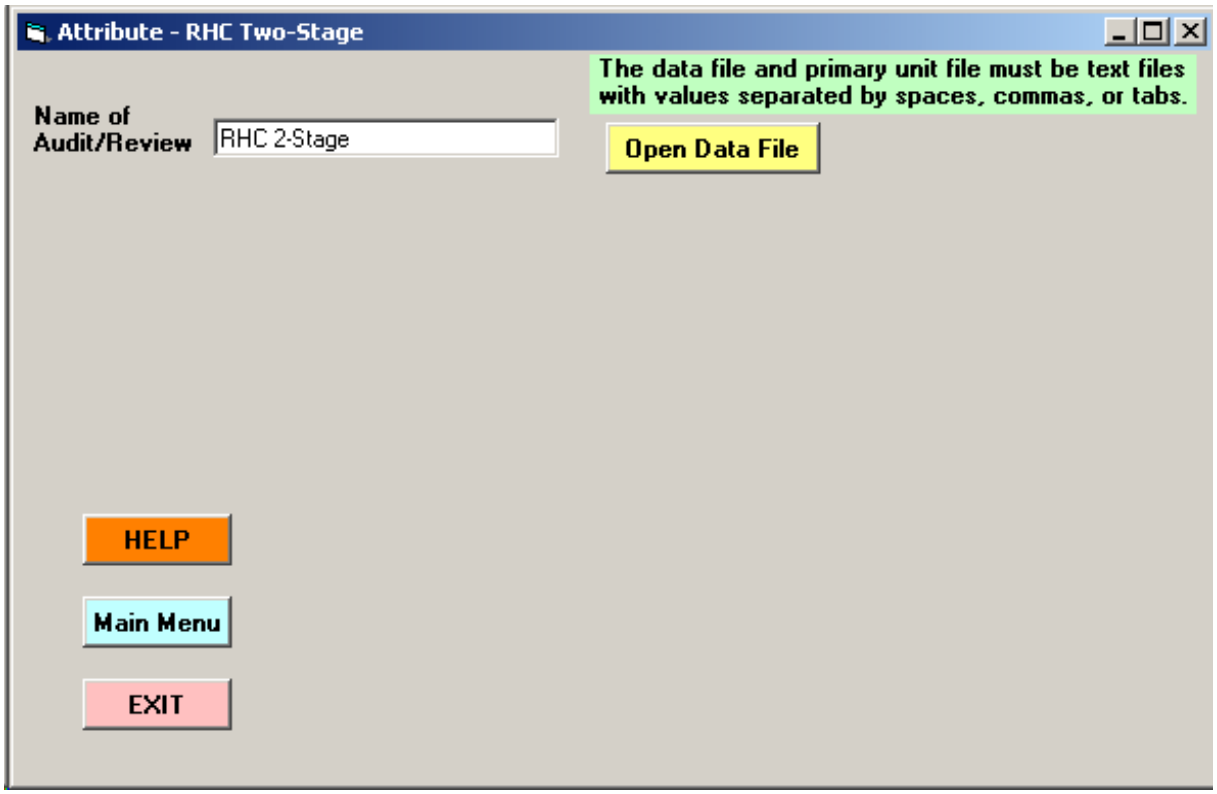
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

RHC TWO STAGE

Purpose

This program performs a two-stage attribute appraisal using the RHC methodology. This multistage procedure determines a confidence interval that uses approximate probability proportional to size (pps) sampling whereby the relative sizes of the sampling units are considered when selecting primary units to include in the sample.

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.

Prior to executing this program, the user must create a data file that contains identifying data and two pieces of information for each secondary unit sampled. The first piece of information is the sample size for this secondary unit. The second value in this line is the number of sample items that met the criteria established for the sample review. These two pieces of information should be separated by one or more spaces, commas, or tabs. Commas should not be used within numeric values (e.g., use 2500 rather than 2,500).

The data file containing the above information must be stored in a text file format. There are several ways the users 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).

7483 200 42

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.

200 - This is the sample size (number of examined units) for this secondary unit.

42 - This is the number of sample items that met the criteria established for the sample review.

The file used in this illustration will be C:\TEMP\RHC2DATA.TXT. After clicking on **Open Data File**, locate this file using the standard Windows "Open" file screen and click on **Open**.

When sample items are selected using the RHC Sample Selection program, a file is created containing information on the clustering of the primary units in the universe as well as the

primary units selected. The **Open Primary Unit File** query is requesting the file created from the RHC Sample Selection program (e.g., C:\TEMP\RHC2PU.TXT).

If the user cannot find this file, a file containing the following data must be created before the appraisal can be performed. The file must be stored in a text file format. The 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).

PRIMARY #1 1500 20100 95500 6

Explanation:

PRIMARY #1 - This is a description of a primary unit in the universe. A maximum of 30 characters, including spaces, will be accepted.

1500 - This is the quantity of secondary units in the particular primary unit.

20100 - This is the size factor assigned by the user to the primary unit (e.g., number of transactions or square footage). A decimal point may be included in this value.

95500 - This is the total size factor for the group in which the primary unit was randomly placed. A decimal point may be included in this value.

6 - This is the quantity of primary units that were randomly placed into this group.

The above entries for each primary unit must be placed on the same line of text. Each line of text must only include the entries for one primary unit. Values within a line can be separated by commas, by one or more spaces, or by using the tab key. Commas should not be used within numeric values (e.g., use 2500 rather than 2,500).

After clicking on **Open Primary Unit File**, locate this file using the standard Windows "Open" file screen and click on **Open**. The screen immediately following the data files will appear.

The data file used in this illustration contains data for 27 universities having one or more state-supported research grants. The universe consists of all charge vouchers recorded for these grants and the purpose of the audit is to estimate the proportion of vouchers containing improper charges. Three universities (UNIV5, UNIV20, and UNIV22) were selected using the RHC Sample Selection program. For each selected university, 250 vouchers were audited. The sample data file and primary unit file are shown next.

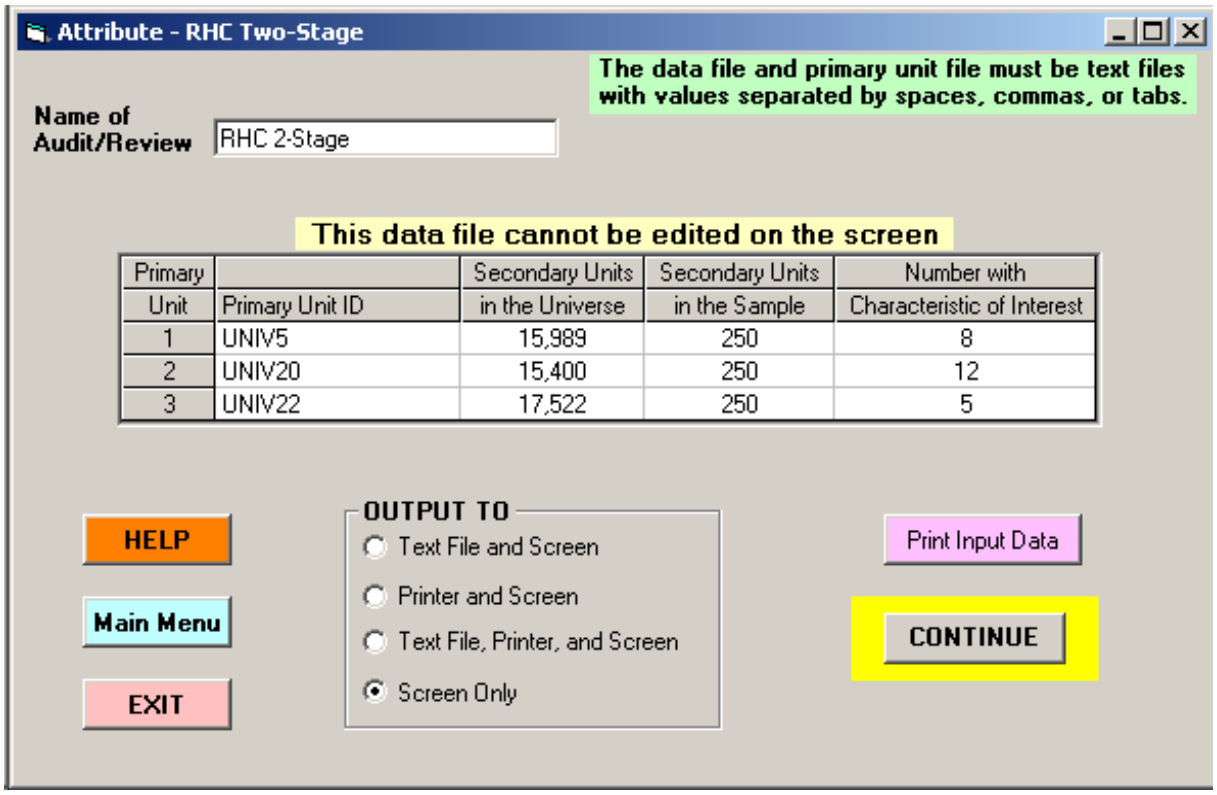
Sample data file C:\TEMP\RHC2DATA.TXT

```
1    250   8
2    250  12
3    250   5
```

Primary unit file C:\TEMP\RHC2PU.TXT

```
UNIV5          15989          11          67          9
UNIV20         15400          10          93          9
UNIV22         17522          13          93          9
```

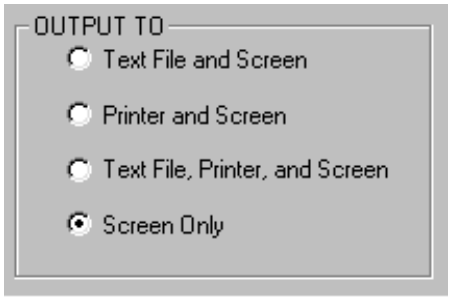
> **NOTE**: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.



The user cannot edit any of the values in the grid. To print this data set, click on **Print Input Data**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. By selecting the appropriate printer and clicking on **OK**, the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The output will first display the data entered by the user for each of the primary units. Then the output will display the results of the appraisal. The following will be displayed in the results portion.

P.U. NBR	For conserving space, the sample primary unit from each group is assigned the group number. For example, the primary unit sampled from group #1 is assigned the number 1 for identification purposes.

PRIMARY UNIT ID	The description assigned by the user to each primary unit.
SECONDARY UNIVERSE	The number of secondary units in the population of a particular primary unit.
PRIMARY UNIT SIZE	The size value selected by the user for assigning a weight to each primary unit. The weight factor could be, for example, number of employees, volume of transactions, or square footage.
GROUP SIZE	The summation of the size values for all the primary units in the particular group.
UNITS IN GROUP	The number of primary units that were placed randomly in a particular group.

For each sampled primary unit, the following information is displayed.

SAMPLE SIZE	The number of sampled secondary units contained in this particular sampled primary unit.
SAMPLE MEAN	The proportion of secondary units containing the characteristic of interest. It is obtained by dividing the number of items in the sample containing the characteristic of interest by the number of sample items.
SECONDARY UNIVERSE	The number of secondary units in the universe for this particular sampled primary unit. This value was originally supplied by the user in the file containing the Primary Unit information.
SIZES RATIO	The ratio of the size of the group containing this particular primary unit to the size of the primary unit itself.
POINT ESTIMATE	The estimate of the universe total for the ***group*** of primary units containing this particular primary unit. For example, suppose the primary units are split into 10 random groups, each containing 5 primary units. Suppose further that the primary unit under discussion lies in group #8. The POINT ESTIMATE refers to the estimate of the universe total of the five primary units in group #8. This would be repeated for the remaining sampled primary units. The point estimates are then totaled to obtain the estimate of the universe total.

--- VARIANCE COMPONENTS —

WITHIN VARIANCE	In the derivation of the standard error, the contribution of the variability of the secondary units.
BETWEEN VARIANCE	In the derivation of the standard error, the contribution of the variability of the primary units.
TOTAL VARIANCE	The sum of WITHIN VARIANCE and BETWEEN VARIANCE. The square root of this value is the STANDARD ERROR.

The results of the above information are then used in the final overall projection. The following information appears:

PRIMARY UNITS SAMPLED	The number of primary units in the sample.
PRIMARY UNITS NOT SAMPLED	The number of primary units in the population minus the number of primary units in the sample.
PRIMARY UNITS IN POPULATION	The total number of primary units in the population.
PROJECTED QUANTITY IN UNIVERSE	The point estimate of population total. A single estimate for a universe value based on the summation of the point estimates for each group.
STANDARD ERROR	A measurement of the standard deviation of the estimate for the population total. It is this value that determines the width of the corresponding confidence intervals.
CONFIDENCE LEVEL	This indicates the confidence (80%, 90%, or 95%) that the user has that the actual population total will fall within the corresponding confidence interval.
LOWER LIMIT	The lower bound of the confidence interval. It is based on subtracting the precision amount from the point estimate.

UPPER LIMIT	The upper bound of the confidence interval. It is based on adding the precision amount to the point estimate.
PRECISION AMOUNT	A measurement of the closeness of the sample estimate and the corresponding population value. For a 90% confidence interval, the user would be 90% confident that the estimated population total (PROJECTED QUANTITY IN UNIVERSE) would be within this amount of the actual value. The precision amount is calculated by multiplying the standard error by the appropriate Z value (Z-VALUE USED).
PRECISION PERCENT	The result of dividing the precision amount by the point estimate.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

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 below. The printer output is identical.

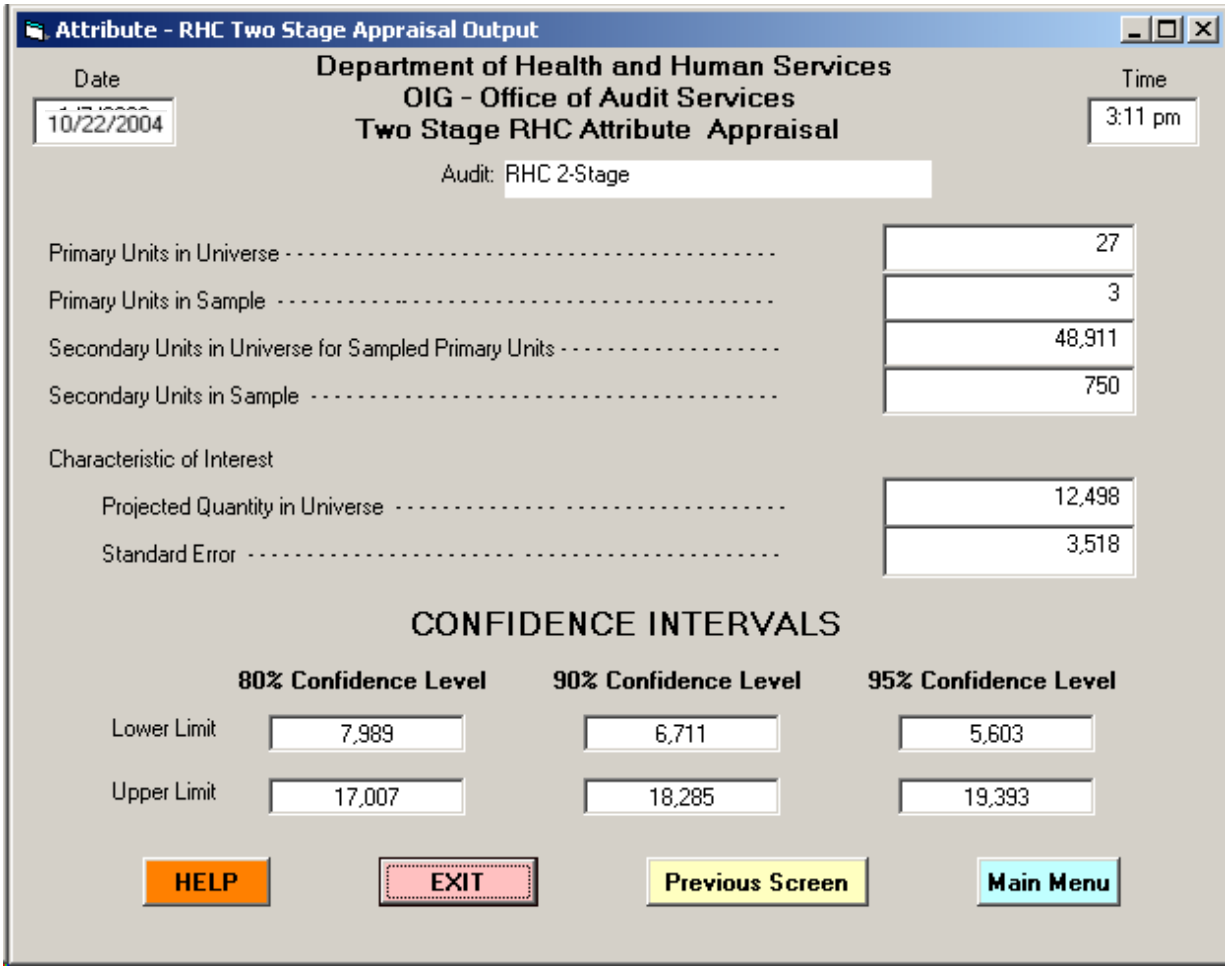
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004          TWO STAGE RHC ATTRIBUTE APPRAISAL       Time: 15:11
                          AUDIT/REVIEW: RHC 2-Stage

                  NAME OF DATA FILE: C:\TEMP\RHC2DATA.TXT
                NAME OF PRIMARY UNIT FILE: C:\TEMP\RHC2PU.TXT
                       OUTPUT FILE: C:\TEMP\OUTDISK.txt

             PRIMARY  SAMPLE  == ATTRIBUTE ==
              UNIT     SIZE     SAMPLE TOTAL
             =======  ======   =============
                1       250              8
                2       250             12
                3       250              5

             TOTALS     750             25

                                                                      UNITS
P.U.                              SECONDARY      PRIMARY                IN
NBR  PRIMARY UNIT ID               UNIVERSE     UNIT SIZE     GROUP SIZE  GROUP
==== ========================== ============= ============= ============= =====
  1  UNIV5                            15,989            11            67     9
  2  UNIV20                           15,400            10            93     9
  3  UNIV22                           17,522            13            93     9

TOTALS:                             48,911            34           253    27

   P.U.  SAMPLE                    SECONDARY      SIZES       POINT
   NBR    SIZE     SAMPLE MEAN      UNIVERSE      RATIO     ESTIMATE
   ====  ======  ==============  =============  ========  =============
     1      250             .03         15,989     6.091       3,116.40
     2      250             .05         15,400     9.300       6,874.56
     3      250             .02         17,522     7.154       2,506.99

TOTALS:     750                         48,911                12,497.96

                         --- VARIANCE COMPONENTS ---

   P.U.             WITHIN                BETWEEN               TOTAL
   NBR             VARIANCE              VARIANCE             VARIANCE
   ====            ========              ========             ========
     1           190,680.61             62,730.47            253,411.08
     2           398,194.39          6,287,762.71          6,685,957.11
     3           170,420.81          5,266,818.04          5,437,238.85

TOTALS:          759,295.82         11,617,311.23         12,376,607.04


PRIMARY UNITS SAMPLED:                                              3
PRIMARY UNITS NOT SAMPLED:                                         24
PRIMARY UNITS IN POPULATION:                                       27
```

```
PROJECTED QUANTITY IN UNIVERSE:                                  12,498
STANDARD ERROR:                                                   3,518

CONFIDENCE LEVEL              80 PERCENT        90 PERCENT        95 PERCENT
LOWER LIMIT                        7,989             6,711             5,603
UPPER LIMIT                       17,007            18,285            19,393
PRECISION AMOUNT                   4,509             5,787             6,895
PRECISION PERCENT                 36.07%            46.30%            55.17%
Z-VALUE USED              1.281551565545    1.644853626951    1.959963984540
```

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

Output to Screen

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



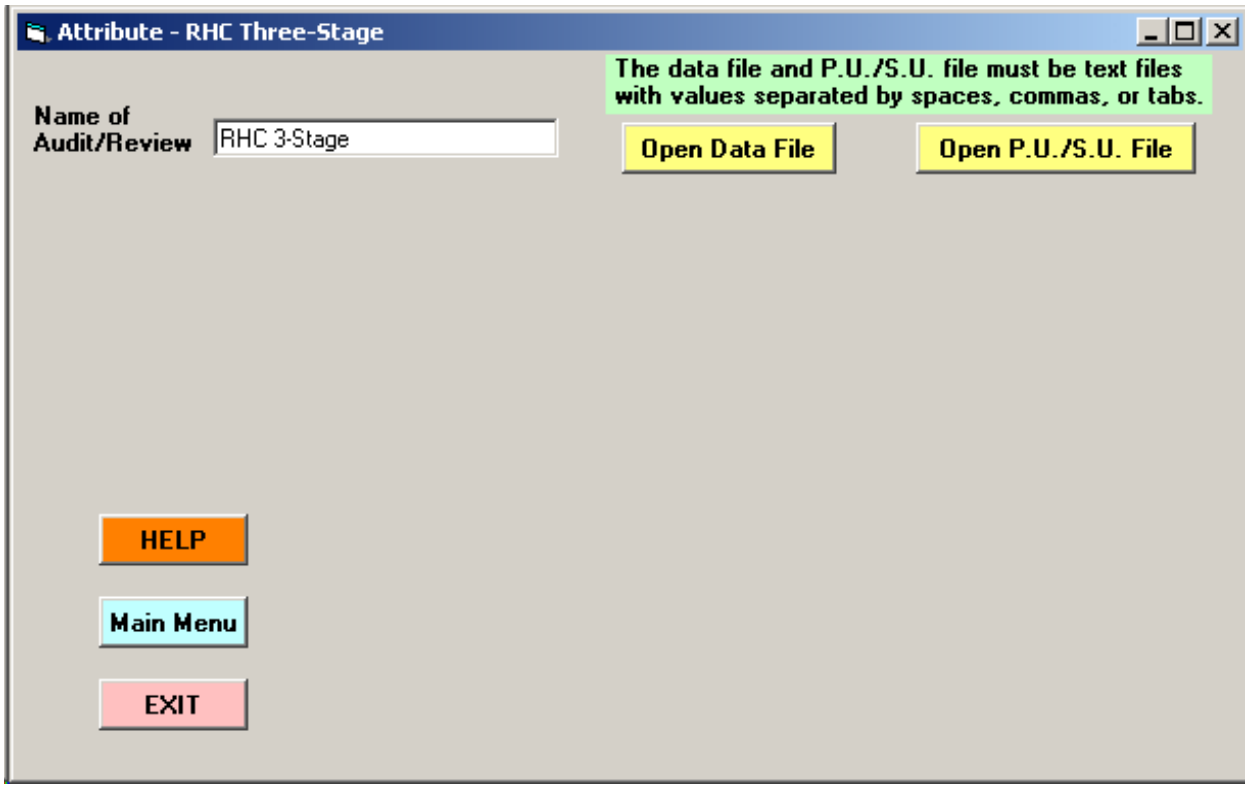
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

RHC THREE STAGE

Purpose

This program performs a three-stage attribute appraisal using the RHC methodology. This multistage procedure determines a confidence interval that uses approximate probability proportional to size (pps) sampling whereby the relative sizes of the sampling units are considered when selecting primary and secondary units to include in the sample.

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.

Prior to executing this program, the user must create a data file that contains identifying data and two pieces of information for each third-stage item sampled. The first piece of information is the sample size for this third-stage unit. The second value in this line is the number of sample items that met the criteria established for the sample review. These two pieces of information should be separated by one or more spaces, commas, or tabs.

The data file containing the above information must be stored in a text file format. There are several ways the users 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).

4.12 180 10

Explanation:

4.12 - This identifies the primary unit number (4, here) and the secondary unit number (12, here). The primary unit numbers should be numbered sequentially (1, 2, 3, ...) and the secondary unit numbers within a sampled primary unit should also be numbered sequentially.

180 - This is the number of third-stage units sampled within this particular primary unit/secondary unit.

10 - This is the number of sample items that met the criteria established for the sample review.

The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (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.

The file used in this illustration will be C:\TEMP\RHC3DATA.TXT. After clicking on **Open Data File**, locate this file using the standard Windows "Open" file screen and click on **Open**.

When sample items are selected using the RHC Sample Selection program, a file was created containing information on the clustering of the primary units/secondary units in the universe as well as the primary units/secondary units selected. The **Open P.U./S.U. File** query is requesting the file created from the RHC Sample Selection Program. The primary unit information is entered first, followed by the information for the secondary units. The format is as follows:

PRIMARY #1	**20**	**9000**	**30600**	**8**	**3**
SECONDARY #1	**3500**	**1000**	**4400**	**10**	

Explanation:

PRIMARY #1 - This is a description of the sampled primary unit. The description should be no longer than 30 characters in length.

20 - This is the quantity of secondary units in this primary unit.

9000 - This is the size factor assigned by the user to the primary unit. A decimal point may be included in this value.

30600 - This is the size factor for the group from which the primary unit was selected. A decimal point may be included in this value.

8 - This is the number of primary units in the group from which this primary unit was selected.

3 - This is the number of secondary units sampled from this primary unit for this appraisal.

SECONDARY #1 - This is a description of the sampled secondary unit. A maximum of 30 characters, including spaces, will be accepted.

3500 - This is the quantity of third-stage units in this secondary unit. The entry may contain a decimal point.

1000 - This is the secondary size factor used in weighting the secondary unit. A decimal point may be incorporated in the number.

4400 - This is the size factor for the secondary group that this secondary item was sampled from. A decimal point may be used in the value.

10 - This is the number of secondary units in this group from which this secondary unit was selected.

The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (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.

After clicking on **Open P.U. / S.U. File**, locate this file within the standard Open File window and click on **Open**. Next, click on **READ FILES** and the screen immediately following the data files will appear.

The data file used in this illustration contains data for three U.S. regions (primary units) and universities (secondary units) within each region having one or more state-supported research grants. The universe consists of all charge vouchers (third-stage units) recorded for these grants and the purpose of the audit is to estimate the proportion of vouchers containing improper charges. Four regions (REGION3, REGION7, REGION8, REGION9) were selected from a universe of 12 using the RHC Sample Selection program. Using the RHC Sample Selection program on each region, 10 universities are selected. For each selected university, approximately 20% of the vouchers were audited. The data file and primary unit/secondary unit file immediately follow.

Sample Data file C:\TEMP\RHC3DATA.TXT

```
1.1     7   2   ⬅ Data for Region 3
1.2    15   4
1.3    15   3
1.4    12   2
1.5    12   5
1.6     6   2
1.7    11   2
1.8     8   4
1.9     5   1
1.10   11   3
2.1    11   2   ⬅ Data for Region 7
2.2    11   4
2.3    13   2
```

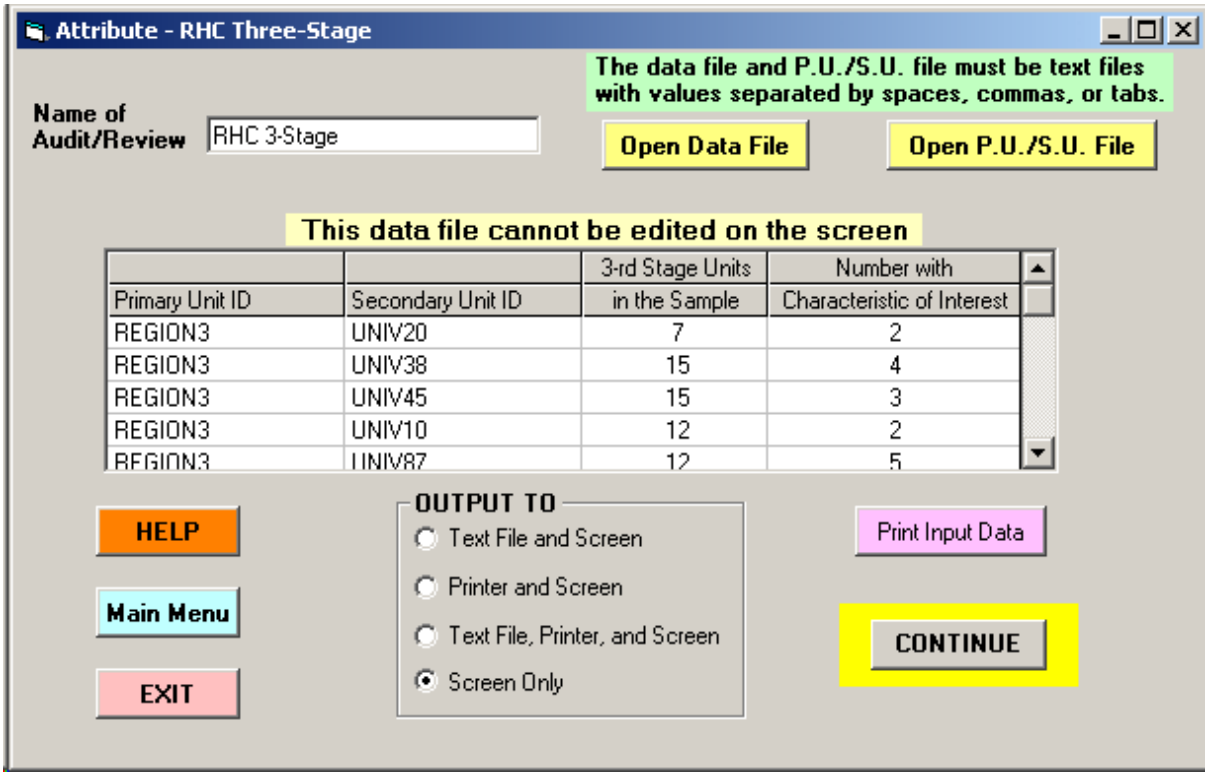
```
2.4   16   6
2.5   13   3
2.6    6   2
2.7    8   3
2.8   11   2
2.9   13   2
2.10  12   3
3.1    7   1   ⬅ Data for Region 8
3.2   15   6
3.3   10   2
3.4    8   4
3.5   14   3
3.6   10   3
3.7    5   0
3.8    5   1
3.9   13   0
3.10  15   3
4.1   13   5   ⬅ Data for Region 9
4.2   15   6
4.3   14   1
4.4   14   3
4.5   11   3
4.6   15   4
4.7   10   0
4.8    7   2
4.9    5   2
4.10  11   5
```

Primary unit file C:\TEMP\RHC3PUSU.TXT

```
REGION3                    91        720       3280          3     10
UNIV20                     37          6         73          9
UNIV38                     74         11         70          9
UNIV45                     73         11         82          9
UNIV10                     60          9         81          9
UNIV87                     62         10         59          9
UNIV82                     30          5         68          9
UNIV60                     54          9         75          9
UNIV69                     39          7         76          9
UNIV34                     26          4         60          9
UNIV54                     57          9         76         10
REGION7                   102        960       2210          3     10
UNIV1                      56         10         89         10
UNIV60                     56         10         96         10
UNIV59                     67         13         94         10
UNIV99                     80         14         91         10
UNIV85                     67         13         93         10
UNIV37                     31          6        103         10
UNIV34                     42          8        106         10
UNIV16                     53         10         83         10
UNIV12                     66         13         90         11
UNIV52                     60         11        115         11
REGION8                   118       1300       3710          3     10
```

UNIV19	34	8	137	11	
UNIV104	77	16	122	11	
UNIV66	49	11	122	12	
UNIV110	38	9	137	12	
UNIV83	70	15	121	12	
UNIV14	48	10	147	12	
UNIV78	27	7	122	12	
UNIV105	27	6	144	12	
UNIV12	65	14	122	12	
UNIV112	75	16	126	12	
REGION9	122	1320	2800	3	10
UNIV37	64	14	147	12	
UNIV92	73	15	125	12	
UNIV47	71	15	130	12	
UNIV54	70	15	131	12	
UNIV97	56	12	138	12	
UNIV66	76	16	122	12	
UNIV116	50	10	140	12	
UNIV29	33	8	128	12	
UNIV18	26	7	132	13	
UNIV107	55	11	127	13	

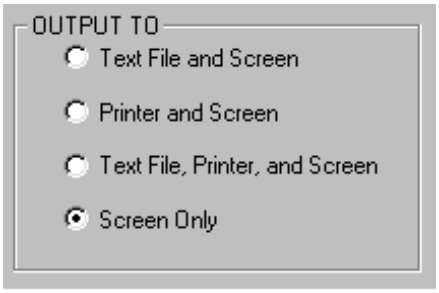
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.



The user cannot edit any of the values in the grid. To print this data set, click on **Print Input Data**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen, or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. By selecting the appropriate printer and clicking on **OK**, the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The output will first display the data entered by the user for each of the primary units. Then the output will display the results of the appraisal. The following will be displayed in the results portion:

--- POINT ESTIMATES ---

PRIMARY/ SECONDARY IDENTIFICATION	The description information obtained from the file of universe data identified by the user.
SAMPLE MEAN	The proportion of third-stage items having the characteristic of interest. It is obtained by dividing the number of items in the sample having the characteristic of interest by the number of items in the sample.
SIZES RATIO	The ratio of the size of the group containing this particular secondary unit to the size of the secondary unit itself. The size factors are obtained from the universe file indicated by the user.
POINT ESTIMATE	The estimate of the universe total for the group of secondary units containing this particular secondary unit. For example, suppose the secondary units are split into 10 random groups, each containing 5 secondary units. Suppose further that the sampled secondary unit lies in group #8. The point estimate refers to the estimate of the universe total of the five secondary units in group #8. This would

be repeated for the remaining sampled secondary units.

--- VARIANCE COMPONENTS FOR PRIMARY UNITS ---

WITHIN VARIANCE	For each sampled primary unit, the contribution of the third-stage variation.
BETWEEN VARIANCE	For each sampled primary unit, the contribution of the second-stage variation.
TOTAL VARIANCE	For each sampled primary unit, the sum of WITHIN VARIANCE and BETWEEN VARIANCE. This value represents the total variation obtained by applying a two-stage RHC procedure to the sampled primary unit.

--- COMBINED VARIANCE COMPONENTS ---

STAGE 1	In the derivation of the standard error, the contribution of the first-stage (primary) units.
STAGES 2 AND 3	In the derivation of the standard error, the contribution of the second-stage (secondary) and third-stage units.
TOTAL VARIANCE	The sum of the values for STAGE 1 and STAGES 2 AND 3. The square root of this value is the STANDARD ERROR.

SUMMARY SECTION: The results of the above information are then used in the final overall projection. The following information appears:

--- SUMMARY OF APPRAISAL RESULTS ---

PRIMARY UNITS SAMPLED	The quantity of primary units selected in this sample.
PRIMARY UNITS NOT SAMPLED	The number of primary units in the population minus the number of primary units in the sample.

TOTAL PRIMARY UNITS	The total number of primary units in the population.
PROJECTED QUANTITY IN UNIVERSE	The overall point estimate is a single estimate for a universe value based on each primary unit point estimate multiplied by (A/B) where A is the size of the group containing the primary unit, B is the size of the primary unit, and the products are summed over all primary units.
OVERALL STANDARD ERROR	A measurement of the standard deviation of the estimate for the population total. It is this value that determines the width of the corresponding confidence intervals.
CONFIDENCE LEVEL	This indicates the confidence the user has that the actual population total will fall within the corresponding confidence interval.
LOWER LIMIT	The lower bound of the confidence interval. It is based on subtracting the precision amount from the point estimate.
UPPER LIMIT	The upper bound of the confidence interval. It is based on adding the precision amount to the point estimate.
PRECISION AMOUNT	A measurement of the closeness of the sample estimate and the corresponding population value. For a 90% confidence interval, the user would be 90% confident that the estimated population total (OVERALL POINT ESTIMATE) would be within this amount of the actual value. The precision amount is calculated by multiplying the standard error by the appropriate factor ("z" value).
PRECISION PERCENT	The result of dividing the precision amount by the point estimate.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

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 on the next page. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004      THREE STAGE RHC ATTRIBUTE APPRAISAL         Time: 16:02
                          AUDIT/REVIEW: RHC 3-Stage

                     DATA FILE USED: C:\TEMP\RHC3DATA.TXT
          PRIMARY/SECONDARY UNIVERSE FILE USED: C:\TEMP\RHC3PUSU.TXT
                       OUTPUT FILE: C:\TEMP\OUTDISK.txt

                                                   *** ATTRIBUTE ***
     **** SAMPLED UNITS ****        THIRD STAGE  SAMPLE    NO. WITH
PRIMARY / SECONDARY IDENTIFICATION    UNIVERSE    SIZE    ATTRIBUTE
==================================  ===========  ======  ============
REGION3
    UNIV20                                   37       7             2
    UNIV38                                   74      15             4
    UNIV45                                   73      15             3
    UNIV10                                   60      12             2
    UNIV87                                   62      12             5
    UNIV82                                   30       6             2
    UNIV60                                   54      11             2
    UNIV69                                   39       8             4
    UNIV34                                   26       5             1
    UNIV54                                   57      11             3
REGION7
    UNIV1                                    56      11             2
    UNIV60                                   56      11             4
    UNIV59                                   67      13             2
    UNIV99                                   80      16             6
    UNIV85                                   67      13             3
    UNIV37                                   31       6             2
    UNIV34                                   42       8             3
    UNIV16                                   53      11             2
    UNIV12                                   66      13             2
    UNIV52                                   60      12             3
REGION8
    UNIV19                                   34       7             1
    UNIV104                                  77      15             6
    UNIV66                                   49      10             2
    UNIV110                                  38       8             4
    UNIV83                                   70      14             3
    UNIV14                                   48      10             3
    UNIV78                                   27       5             0
    UNIV105                                  27       5             1
    UNIV12                                   65      13             0
    UNIV112                                  75      15             3
REGION9
    UNIV37                                   64      13             5
    UNIV92                                   73      15             6
    UNIV47                                   71      14             1
    UNIV54                                   70      14             3
    UNIV97                                   56      11             3
    UNIV66                                   76      15             4
```

UNIV116	50	10	0
UNIV29	33	7	2
UNIV18	26	5	2
UNIV107	55	11	5
TOTALS	2,440	433	111

--- POINT ESTIMATES ---
*** ATTRIBUTE ***

**** SAMPLED UNITS **** PRIMARY / SECONDARY IDENTIFICATION	SAMPLE MEAN	SIZES RATIO	POINT ESTIMATE
REGION3			
UNIV20	0.29	12.1667	129
UNIV38	0.27	6.3636	126
UNIV45	0.20	7.4545	109
UNIV10	0.17	9.0000	90
UNIV87	0.42	5.9000	152
UNIV82	0.33	13.6000	136
UNIV60	0.18	8.3333	82
UNIV69	0.50	10.8571	212
UNIV34	0.20	15.0000	78
UNIV54	0.27	8.4444	131
TOTAL			1,244
REGION7			
UNIV1	0.18	8.9000	91
UNIV60	0.36	9.6000	195
UNIV59	0.15	7.2308	75
UNIV99	0.38	6.5000	195
UNIV85	0.23	7.1538	111
UNIV37	0.33	17.1667	177
UNIV34	0.38	13.2500	209
UNIV16	0.18	8.3000	80
UNIV12	0.15	6.9231	70
UNIV52	0.25	10.4545	157
TOTAL			1,359
REGION8			
UNIV19	0.14	17.1250	83
UNIV104	0.40	7.6250	235
UNIV66	0.20	11.0909	109
UNIV110	0.50	15.2222	289
UNIV83	0.21	8.0667	121
UNIV14	0.30	14.7000	212
UNIV78	0.00	17.4286	0
UNIV105	0.20	24.0000	130
UNIV12	0.00	8.7143	0
UNIV112	0.20	7.8750	118
TOTAL			1,296
REGION9			
UNIV37	0.38	10.5000	258
UNIV92	0.40	8.3333	243
UNIV47	0.07	8.6667	44
UNIV54	0.21	8.7333	131
UNIV97	0.27	11.5000	176
UNIV66	0.27	7.6250	155
UNIV116	0.00	14.0000	0

UNIV29	0.29	16.0000	151
UNIV18	0.40	18.8571	196
UNIV107	0.45	11.5455	289
TOTAL			1,643

--- VARIANCE COMPONENTS FOR PRIMARY UNITS ---

**** SAMPLED UNITS **** PRIMARY UNIT IDENTIFICATION	WITHIN VARIANCE	BETWEEN VARIANCE	TOTAL VARIANCE
REGION3	3,898	15,618	19,516
REGION7	4,356	22,749	27,105
REGION8	3,952	75,526	79,478
REGION9	4,749	77,463	82,212

--- COMBINED VARIANCE COMPONENTS ---

STAGE 1	STAGES 2 AND 3	TOTAL VARIANCE
2,580,562	552,511	3,133,073

*** ATTRIBUTE ***

--- SUMMARY OF APPRAISAL RESULTS ---

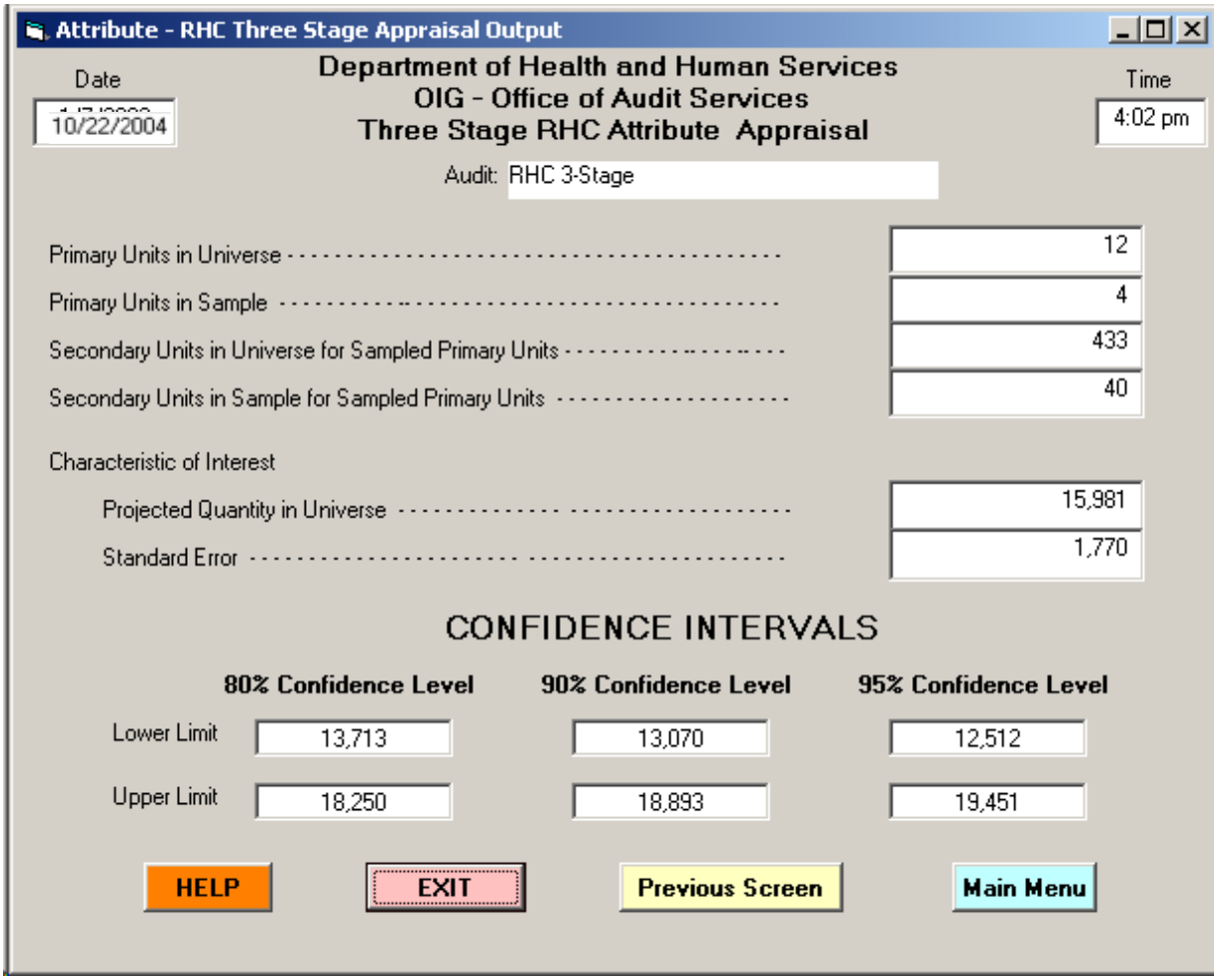
PRIMARY UNITS SAMPLED	4
PRIMARY UNITS NOT SAMPLED	8
TOTAL PRIMARY UNITS	12
PROJECTED QUANTITY IN UNIVERSE	15,981
STANDARD ERROR	1,770

CONFIDENCE LEVEL	80 PERCENT	90 PERCENT	95 PERCENT
LOWER LIMIT	13,713	13,070	12,512
UPPER LIMIT	18,250	18,893	19,451
PRECISION AMOUNT	2,268	2,911	3,469
PRECISION PERCENT	14.19%	18.22%	21.71%
Z-VALUE USED	1.281551565545	1.644853626951	1.959963984540

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

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:



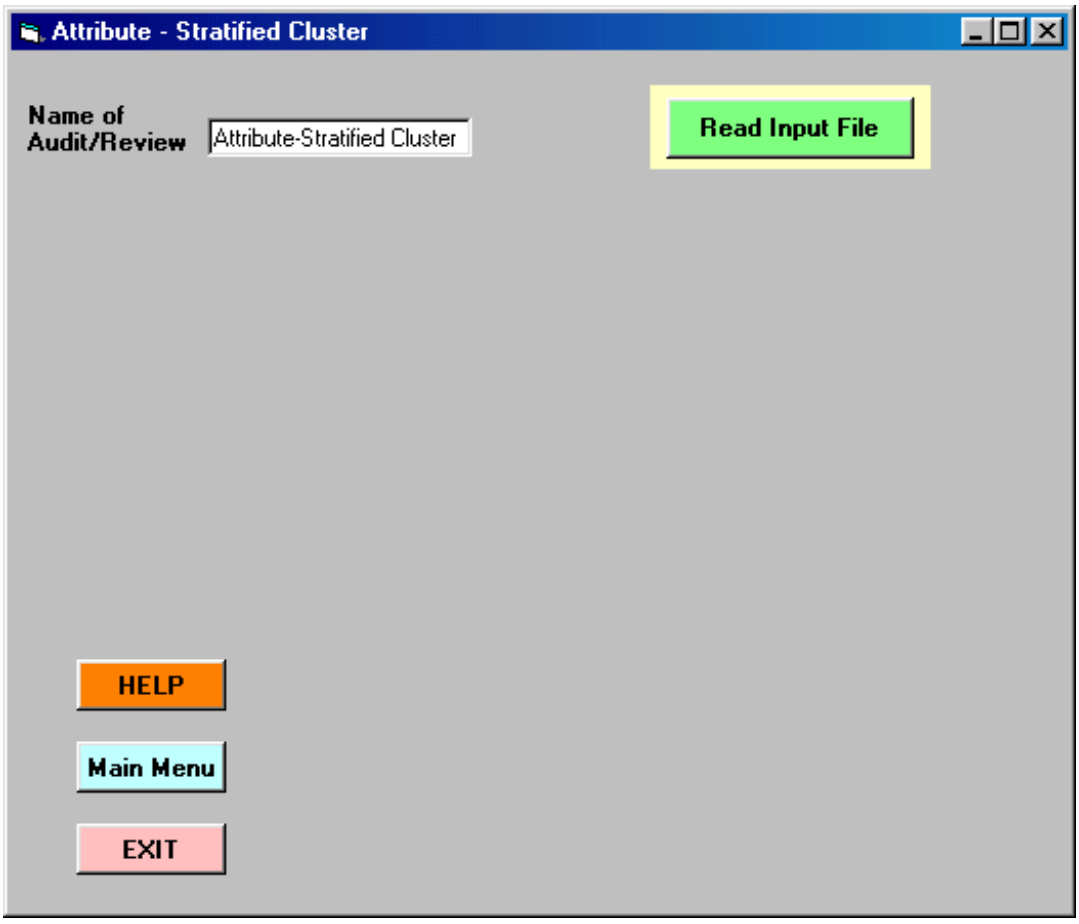
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

STRATIFIED CLUSTER

Purpose

This program computes the point estimate, standard error, and confidence limits for the universe percentage and universe total number when using a stratified cluster attribute sampling procedure. Attribute sampling is used to determine how frequently an event or type of transaction occurs in a given universe. The results can be reported as a percentage estimate or an estimate of the total number in the universe having this attribute. The program will accept a minimum of one stratum and a maximum of 100 strata. A sample of no more than 150 primary units is allowed within each stratum (i.e., 100 x 150 = 15,000 primary units). This procedure is used when all secondary items in a sampled primary unit are reviewed for sampling purposes.

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.

Read input file

When the name of review/audit has been specified, click on the **Read Input File** button.

Prior to executing this program, the user *must* create a data file that contains the results of the stratified cluster attribute sample. The data file contains information about each of the primary and secondary units sampled.

The data file containing the information must 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 storing the data in a spreadsheet software application (e.g., Excel) as a print file. Values within a line can be separated by commas, by one or more spaces, or by using the tab key. Commas should not be used within numeric values (e.g., use 2500 rather than 2,500). The data file needs to be formatted in the following manner:

STATE UNIVERSITIES 415 25 2500
UNIV1 8 2

Explanation:

STATE UNIVERSITIES - is the description of a primary unit. The description may be of any length up to 25 characters and include characters, digits, spaces, and punctuation marks.

415 - is the number of primary units in the universe for this stratum.

25 - is the number of primary units in the sample for this stratum.

2,500 - is the number of secondary units in the universe for this stratum.

UNIV1 - is a description of a primary unit that was sampled. The

description may be of any length up to 25 characters and include characters, digits, spaces, and punctuation marks.

8 - is the universe of secondary items in the sampled primary unit. The number also represents the quantity of secondary items being sampled.

2 - is the number of secondary items that met the characteristic of interest to the user.

The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (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 value that contains nonnumeric data.

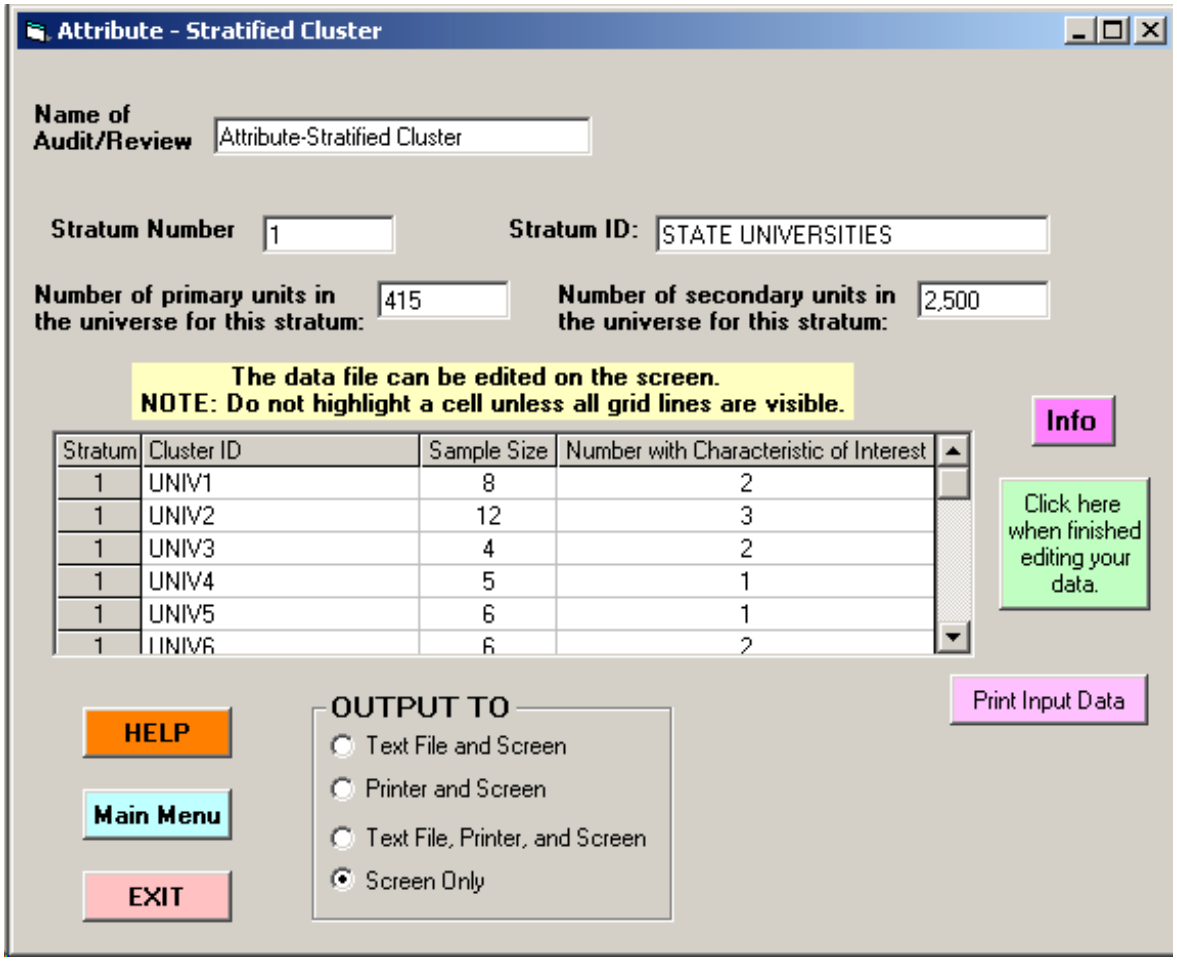
The data file used in this illustration is C:\TEMP\DATACLUS.TXT. The universe consists of 583 universities with health-related research grants. The two strata consist of state universities (415 universities) and private universities (168 universities). Within each stratum, a single-stage cluster sample was obtained by sampling 25 state universities and 10 private universities. The total number of grants in the universe is 2,500 (state universities) and 1,000 (private universities) for a total of 3,500 grants in the entire universe. Of interest is the proportion of grants containing charges after the scheduled completion of the grant.

```
STATE UNIVERSITIES  415  25  2500
UNIV1    8  2
UNIV2   12  3
UNIV3    4  2
UNIV4    5  1
UNIV5    6  1
UNIV6    6  2
UNIV7    7  2
UNIV8    5  2
UNIV9    8  2
UNIV10   3  1
UNIV11   2  0
UNIV12   6  2
UNIV13   5  1
UNIV14  10  3
UNIV15   9  1
UNIV16   3  1
UNIV17   6  2
UNIV18   5  1
UNIV19   5  1
```

```
UNIV20   4  1
UNIV21   6  1
UNIV22   8  1
UNIV23   7  2
UNIV24   3  1
UNIV25   8  2
PRIVATE UNIVERSITIES  168  10  1000
UNIV1    2  1
UNIV2    5  2
UNIV3    7  2
UNIV4    4  2
UNIV5    3  1
UNIV6    8  3
UNIV7    6  2
UNIV8   10  4
UNIV9    3  1
UNIV10   1  1
```

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

After clicking on **Read Input File** and opening the data file (i.e., C:\TEMP\DATACLUS.TXT), the following screen will appear:

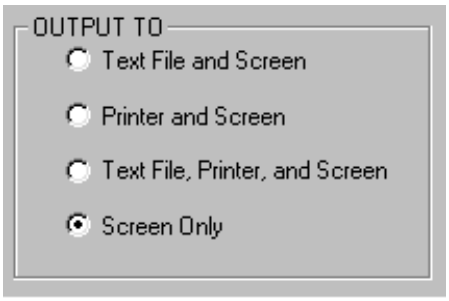


For help in viewing the information for each stratum contained in the four boxes above the grid, click on the **Info** button. If any values within the grid are edited, click on **Click here when finished editing your data** and the **Save Input Data** button will appear.

To save this data set, click on **Save Input 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**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".txt" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. By selecting the appropriate printer and clicking on **OK**, the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The first section of the program output contains the name of the data file along with the contents of this file in summary form. In addition, the output displays the percent meeting the criteria and projected quantity for each stratum.

The next section of output contains overall strata totals in terms of universes, sample sizes and quantity of items with the characteristic of interest. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following information is also presented:

OVERALL TOTAL (%)	The estimate of the universe proportion having the characteristic of interest.

OVERALL TOTAL (#)	The estimate, expressed in terms of secondary units, of items having the characteristic of interest.
OVERALL STANDARD ERROR (%)	An estimate of the standard deviation of the proportion estimate; a value which determines the width of the corresponding confidence intervals for this parameter.
OVERALL STANDARD ERROR (#)	An estimate of the standard deviation of the universe total estimate; a value which determines the width of the corresponding confidence intervals for this parameter.
CONFIDENCE LEVEL	The confidence (80%, 90%, or 95%) that the actual proportion (or total number in the universe) will fall within the corresponding confidence interval.
LOWER LIMIT FOR TOTAL	The lower limit of the 80%, 90%, and 95% confidence interval for the universe total number.
UPPER LIMIT FOR TOTAL	The upper limit of the 80%, 90%, and 95% confidence interval for the universe total number.
LOWER LIMIT FOR PROPORTION	The lower limit of the 80%, 90%, and 95% confidence interval for the universe proportion.
UPPER LIMIT FOR PROPORTION	The upper limit of the 80%, 90%, and 95% confidence interval for the universe proportion.

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 immediately following. The printer output is identical but does not include the output file name if the output was not also saved to a text file.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                       OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004    STRATIFIED CLUSTER ATTRIBUTE APPRAISAL        Time: 14:35
                AUDIT/REVIEW: Attribute - Stratified Cluster
                  NAME OF INPUT FILE: C:\TEMP\DATACLUS.TXT
                     OUTPUT FILE: C:\TEMP\OUTDISK.TXT
```

STRATUM IDENTIFICATION / CLUSTER IDENTIFICATION	SAMPLE UNIVERSE	SAMPLE SIZE	MEETING CRITERIA	PERCENT	PROJECTED QUANTITY
STATE UNIVERSITIES	415	25			
UNIV1	8	8	2		
UNIV2	12	12	3		
UNIV3	4	4	2		
UNIV4	5	5	1		
UNIV5	6	6	1		
UNIV6	6	6	2		
UNIV7	7	7	2		
UNIV8	5	5	2		
UNIV9	8	8	2		
UNIV10	3	3	1		
UNIV11	2	2	0		
UNIV12	6	6	2		
UNIV13	5	5	1		
UNIV14	10	10	3		
UNIV15	9	9	1		
UNIV16	3	3	1		
UNIV17	6	6	2		
UNIV18	5	5	1		
UNIV19	5	5	1		
UNIV20	4	4	1		
UNIV21	6	6	1		
UNIV22	8	8	1		
UNIV23	7	7	2		
UNIV24	3	3	1		
UNIV25	8	8	2		
STRATUM TOTALS	2,500	151	38	25.17%	629

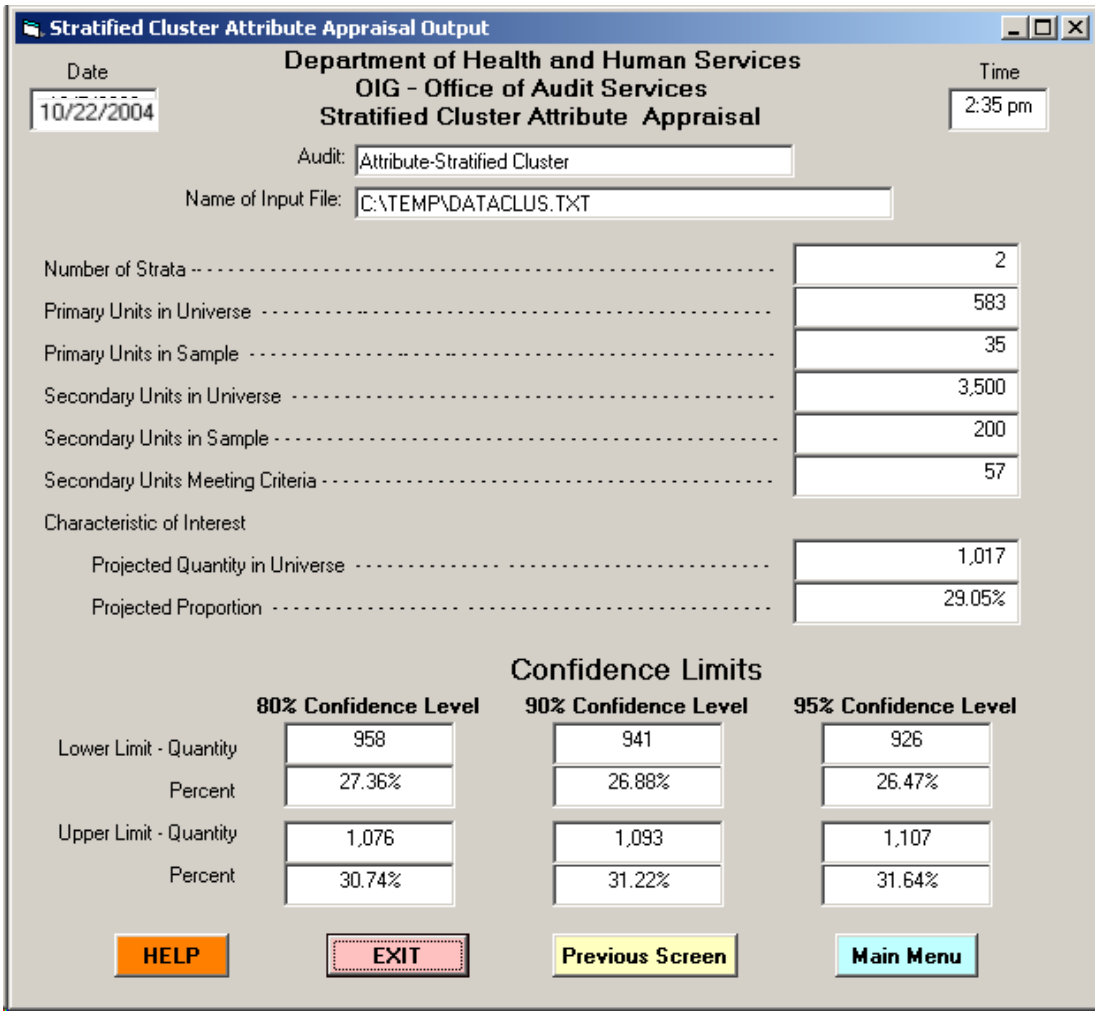
PRIVATE UNIVERSITIES	168	10			
UNIV1	2	2	1		
UNIV2	5	5	2		
UNIV3	7	7	2		
UNIV4	4	4	2		
UNIV5	3	3	1		
UNIV6	8	8	3		
UNIV7	6	6	2		
UNIV8	10	10	4		
UNIV9	3	3	1		
UNIV10	1	1	1		
STRATUM TOTALS	1,000	49	19	38.78%	388
STRATA TOTALS	583	35			
CLUSTER UNIT TOTALS	3,500	200	57		
OVERALL TOTAL				29.05%	1,017
OVERALL STANDARD ERROR				1.32%	46

CONFIDENCE LEVEL	80 PERCENT	90 PERCENT	95 PERCENT
LOWER LIMIT FOR PROPORTION	27.36%	26.88%	26.47%
UPPER LIMIT FOR PROPORTION	30.74%	31.22%	31.64%
LOWER LIMIT FOR TOTAL	958	941	926
UPPER LIMIT FOR TOTAL	1,076	1,093	1,107

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

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:



Stratified Cluster Attribute Appraisal Output

Department of Health and Human Services
OIG - Office of Audit Services
Stratified Cluster Attribute Appraisal

Date: 10/22/2004
Time: 2:35 pm

Audit: Attribute-Stratified Cluster
Name of Input File: C:\TEMP\DATACLUS.TXT

Number of Strata	2
Primary Units in Universe	583
Primary Units in Sample	35
Secondary Units in Universe	3,500
Secondary Units in Sample	200
Secondary Units Meeting Criteria	57
Characteristic of Interest	
Projected Quantity in Universe	1,017
Projected Proportion	29.05%

Confidence Limits

	80% Confidence Level	90% Confidence Level	95% Confidence Level
Lower Limit - Quantity	958	941	926
Percent	27.36%	26.88%	26.47%
Upper Limit - Quantity	1,076	1,093	1,107
Percent	30.74%	31.22%	31.64%

HELP EXIT Previous Screen Main Menu

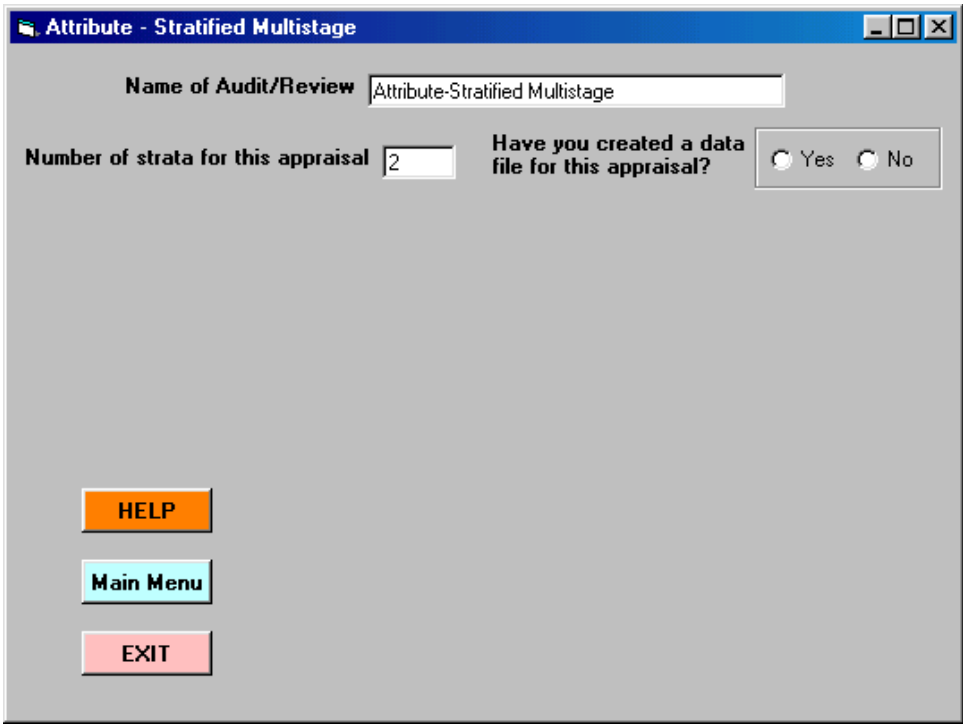
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

STRATIFIED MULTISTAGE

Purpose

This program performs a stratified multistage appraisal based on information gathered from prior multistage appraisals. The user would have initially stratified the clusters (e.g., universities) into two or more categories (e.g., public and private universities). Within each stratum, the user would select a multistage sample. The results of the samples would be appraised using a multistage appraisal program. The point estimate and standard error from each of these appraisals could be placed in a data file or entered interactively by the user.

Input Screen



NOTE: Example is for illustrative purposes only. The sample sizes may not conform to 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.

Number of strata for this appraisal

The user must specify the number of strata that will be appraised. This number must be between 2 and 500.

Have you created a data file?

Prior to executing this program the user can create a data file that contains the point estimate and the standard error for each of the strata. The data file containing the above information must be stored in a text file format. 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, select the input file and click on **Open.**

Once the data file has been created and opened, the values from the file will be displayed on the screen for the user's review. For the data in each stratum, the user needs to use the following format:

16.90 1.99 5600

Explanation:

16.90 - This is the overall estimate for one stratum. This value should be expressed as a percentage (e.g., enter 5 for 5%, not .05).

1.99 - This is the standard error for one stratum. This value should be expressed as a percentage (e.g., enter 5 for 5%, not .05).

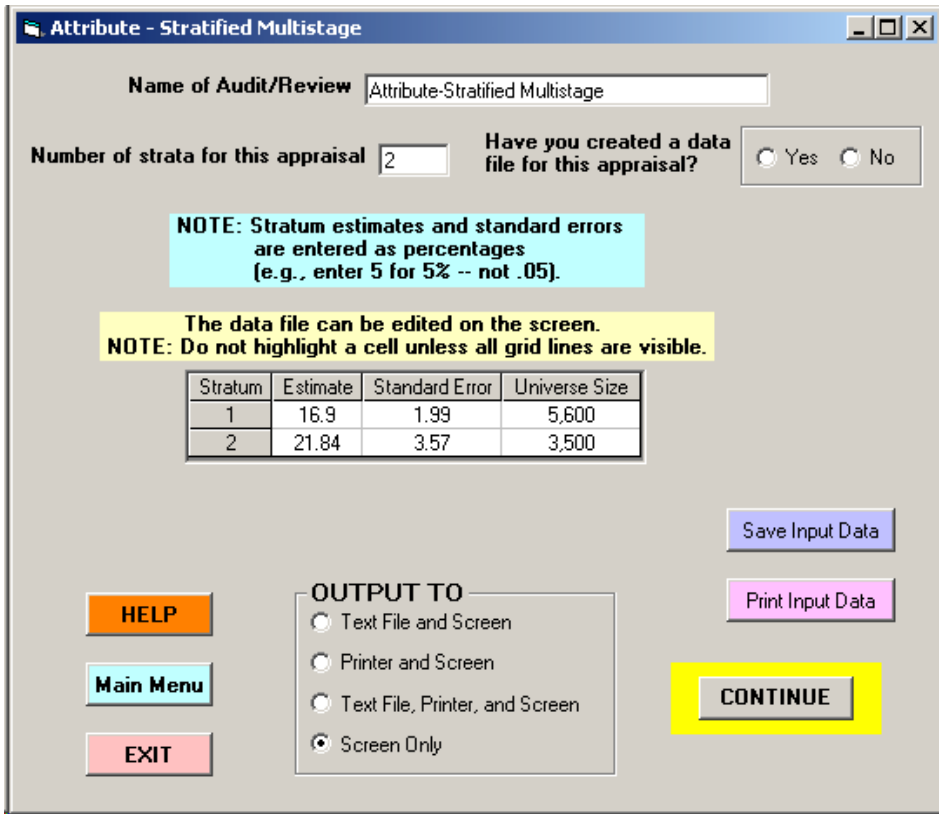
5600 - This is the number of sample units in the universe for the stratum. This is the universe at the most detailed level of the multistage sample. Commas should not be used within this value.

The results for each stratum should be on a separate line. The program assumes a comma, one or more spaces, or a tab as a delimiter between pieces of data. The user needs to enter large numbers (e.g., 10000) without commas (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 value that contains nonnumeric data.

The text file C:\TEMP\DATAMULTI.TXT used in this illustration is shown below. There are two strata consisting of state universities (Stratum 1) and private universities (Stratum 2).

```
16.90   1.99   5600
21.84   3.57   3500
```

After selecting the data file, the following screen will appear:



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

Entering the Sample Results From the Screen

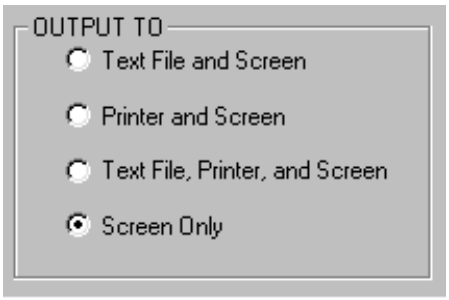
If "No" was selected in response to the query "Have you created a data file for this appraisal?" then a blank grid will appear and the user may enter the data values (e.g., 6) directly using the screen. **NOTE**: This grid structure will not allow the user to use the tab key to move from cell to cell. Each cell must be clicked on before entering its value.

After all values within the grid have been entered, click on the **Click here after entering or editing data** button.

To save this data set, click on **Save Input 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**. Select a printer and click on **OK**.

Output Options

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



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

If the user selects a text file for output, the standard Windows "Save" file screen will appear. Fill in the name of the file in the **File name** box. A text file will be saved with a ".TXT" extension (e.g., C:\TEMP\OUTDISK.TXT). By clicking on the **Save** button, the program will return to the original input screen for this module.

If the user selects the printer for output, the standard Windows "Print" dialog box will appear. Select the appropriate printer and click on **OK**, the program will return to the original input screen for this module.

To continue the appraisal, click on **CONTINUE**.

Program Output

The output will first display the data entered by the user for each of the strata. Then the output will display the results of the appraisal. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels. The following will be displayed in the results portion:

ESTIMATED PERCENTAGE	A single estimate of the universe percentage for all strata of the universe value of the characteristic(s) being measured.
ESTIMATED TOTAL	A single estimate of the universe total for all strata of the universe value of the characteristic(s) being measured.
STANDARD ERROR (PERCENTAGE)	An estimate of the standard deviation of the proportion estimate; a value which determines the width of the corresponding confidence intervals for this parameter.
STANDARD ERROR (TOTAL)	An estimate of the standard deviation of the universe total estimate; a value which determines the width of the corresponding confidence intervals for this parameter.
CONFIDENCE LEVEL	This indicates the confidence that the user has that the actual value of the measured characteristic will fall within the range from the lower to upper limits (confidence interval). The confidence levels are 80%, 90%, and 95%.
LOWER LIMIT FOR PROPORTION	The lower bound of the 80%, 90%, or 95% confidence interval for the universe proportion.
UPPER LIMIT FOR PROPORTION	The upper bound of the 80%, 90%, or 95% confidence interval for the universe proportion.

LOWER LIMIT FOR TOTAL	The lower bound of the 80%, 90%, or 95% confidence interval for the universe total.
UPPER LIMIT FOR TOTAL	The upper bound of the 80%, 90%, or 95% confidence interval for the universe total.

Output to a Text File or Printer

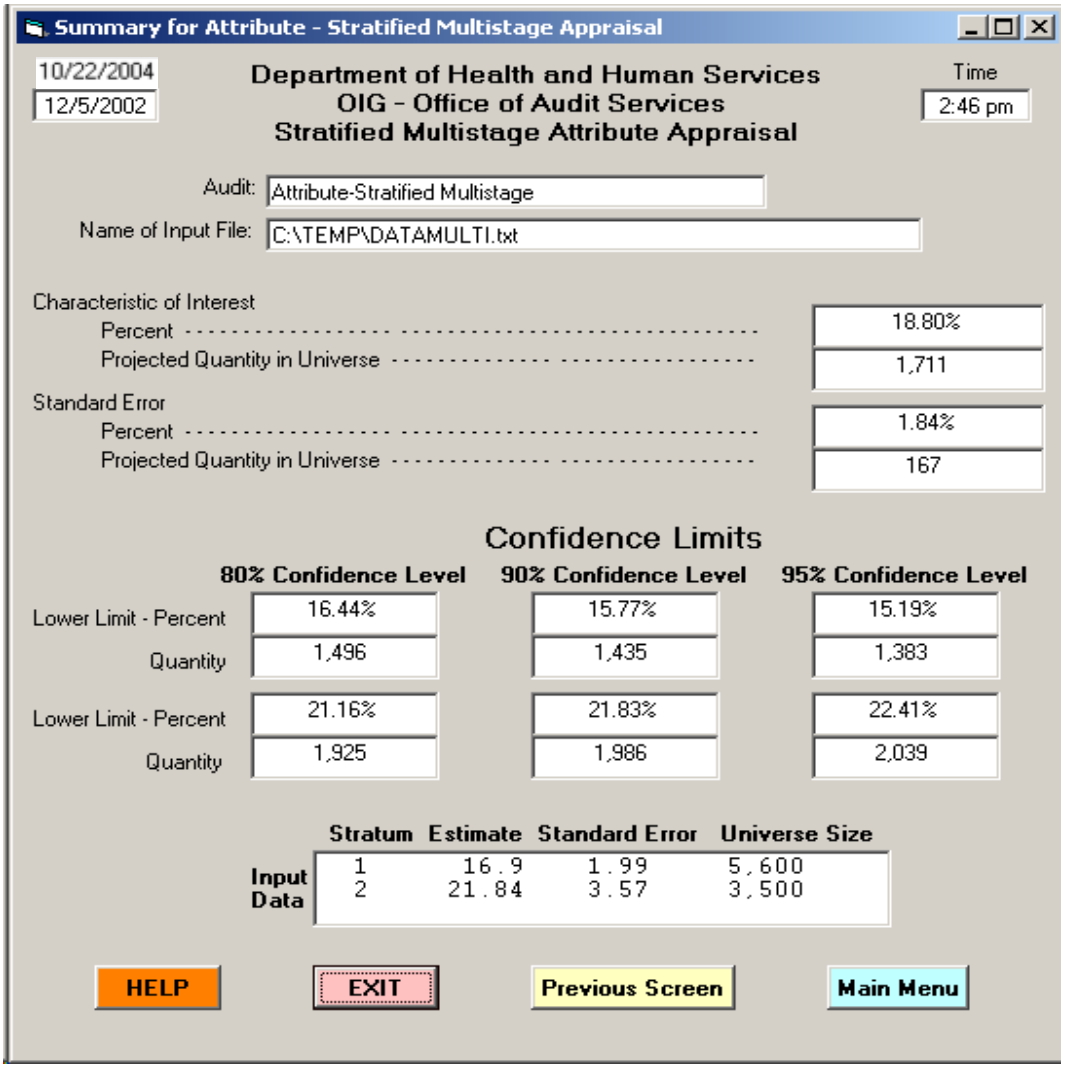
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTDISK.TXT, shown immediately following. The printer output is identical but does not include the output file name if the output was not also saved to a text file.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004     STRATIFIED MULTISTAGE ATTRIBUTE APPRAISAL    Time: 14:46
                AUDIT/REVIEW: Attribute-Stratified Multistage
                       DATA FILE: C:\TEMP\DATAMULTI.TXT
                       OUTPUT FILE: C:\TEMP\OUTDISK.TXT

             THE ESTIMATORS ARE BASED ON THE FOLLOWING ENTRIES:
  STRATUM         ESTIMATE      STANDARD ERROR        UNIVERSE SIZE
      1             16.90%               1.99%                5,600
      2             21.84%               3.57%                3,500


= = = = = = = = = = = = = = = = = = RESULTS = = = = = = = = = = = = = = = = = = =

ESTIMATED PERCENTAGE:              18.80%
ESTIMATED TOTAL:                    1,711

STANDARD ERROR (PERCENTAGE):        1.84%
STANDARD ERROR (TOTAL):               167

CONFIDENCE LEVEL                80 PERCENT        90 PERCENT        95 PERCENT
LOWER LIMIT FOR PROPORTION          16.44%            15.77%            15.19%
UPPER LIMIT FOR PROPORTION          21.16%            21.83%            22.41%

LOWER LIMIT FOR TOTAL                1,496             1,435             1,383
UPPER LIMIT FOR TOTAL                1,925             1,986             2,039
```

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

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:



Summary for Attribute - Stratified Multistage Appraisal

10/22/2004
12/5/2002

Department of Health and Human Services
OIG - Office of Audit Services
Stratified Multistage Attribute Appraisal

Time 2:46 pm

Audit: Attribute-Stratified Multistage

Name of Input File: C:\TEMP\DATAMULTI.txt

Characteristic of Interest	
Percent	18.80%
Projected Quantity in Universe	1,711
Standard Error	
Percent	1.84%
Projected Quantity in Universe	167

Confidence Limits

	80% Confidence Level	90% Confidence Level	95% Confidence Level
Lower Limit - Percent	16.44%	15.77%	15.19%
Quantity	1,496	1,435	1,383
Lower Limit - Percent	21.16%	21.83%	22.41%
Quantity	1,925	1,986	2,039

Input Data

Stratum	Estimate	Standard Error	Universe Size
1	16.9	1.99	5,600
2	21.84	3.57	3,500

HELP EXIT Previous Screen Main Menu

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

Variable Appraisals

OVERVIEW

UNRESTRICTED

STRATIFIED

TWO-STAGE UNRESTRICTED

THREE-STAGE UNRESTRICTED

RHC TWO STAGE

RHC THREE STAGE

STRATIFIED CLUSTER

STRATIFIED MULTISTAGE

POSTSTRATIFICATION

UNKNOWN UNIVERSE SIZE

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.

These appraisal programs provide the correct statistical results only when the proper sample design has been executed. If a stratified sample of inventory cards were drawn from throughout the organization, only the stratified variable appraisal program would generate the proper results. Therefore, the reviewer must be sure at the time the sample approach is developed that the appropriate appraisal methodology will be used.

This package offers the user 10 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.

Two-Stage Unrestricted

The cost of performing a review may affect the sampling methodology used. For example, the inventory could be distributed throughout several locations (primary units) in the United States. A reviewer could randomly select locations and then, from the selected locations, sample inventory cards to be analyzed. This methodology could save travel costs for the reviewer.

Three-Stage Unrestricted

This methodology is similar to the "Two-Stage Unrestricted" with the addition of another level of sampling. Using the inventory example, the reviewer may decide to select by region of the country and then sample locations within the selected regions. Then at the selected locations, the user would select inventory cards.

RHC Two Stage

In certain situations, the reviewer may want to use a multistage sample with greater probability of selecting "larger" units in the universe. For example, a reviewer may want to take an inventory of items at various locations. However, the reviewer may want larger locations to have a greater chance for selection. The Rao, Hartley and Cochran (RHC) methodology allows the reviewer to weight (e.g., by using square footage) the locations (primary units). If this methodology has been used for selecting the sample items, then this module would be used to appraise the sample results.

RHC Three Stage

The appraisal program is similar to the "RHC Two-Stage" with the addition of another level of sampling. For the inventory example, the reviewer may sample by region and then sample locations within the selected regions. The "RHC Sample Selection" must be used to draw the sample.

Stratified Cluster

The selection of a sample item may sometimes be extremely costly in terms of time and resources. However, once the item has been selected it can be reviewed rather quickly. For example, using the inventory cards again, once the reviewer has arrived at the sampled location, it may take only a short amount of time to review all of the inventory items at the location. The reviewer may decide to group the locations by total inventory value (e.g., locations with total inventory over $10 million). For each group (stratum) the reviewer would sample locations. For each selected location, the reviewer would analyze all inventory cards.

Stratified Multistage

This methodology is similar to "Stratified Cluster" with the exception that not all items in the subuniverse are reviewed. In the example discussed above, the assumption was made that the reviewer had sufficient time and resources to analyze all of the inventory cards at the selected locations. This may not be possible. Therefore, this methodology is used when a sample is still needed at each selected location.

Poststratification

The reviewer may wish that the sample had been stratified after evaluating the items. For example, after performing an unrestricted sample of inventory cards, the reviewer may see that certain types of inventory items had a greater chance of discrepancies between the inventory card balance and the actual physical inventory. While post stratification is allowed, if each stratum's universe size is known, the results are less efficient statistically as compared to a stratified sample design.

Unknown Universe Size

The reviewer may have a situation where the size of the universe of items to be sampled is not known and cannot be readily determined. The reviewer could draw a sample from a frame that includes all the desired items. The results from this sample would be used to estimate the universe size. A separate sample would be drawn of review items that meet the reviewer's

criteria as a sample item.

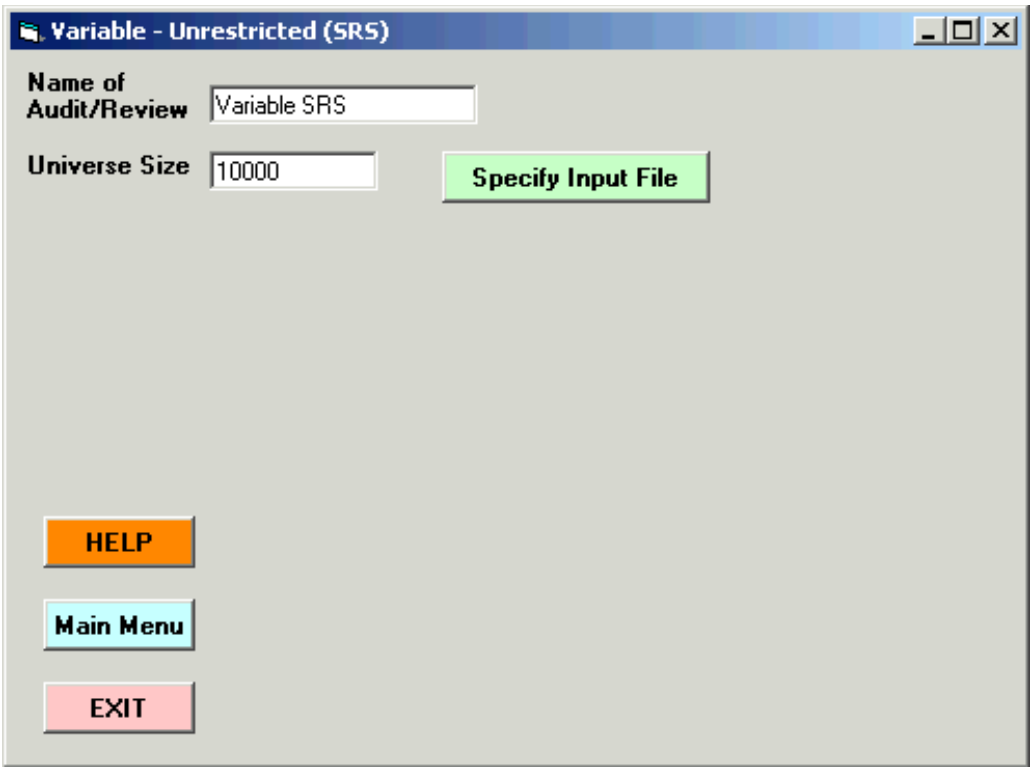
Assume the reviewer wants to sample invoices that are in boxes. The reviewer has no idea how many invoices are in all the boxes. The reviewer could first sample boxes and count the number of invoices in each selected box. An unrestricted variable appraisal would be used to estimate the universe size. The reviewer would then perform another sample to select individual invoices from all the boxes. Another unrestricted variable appraisal would be performed. The results from the two samples would be requested by this program to develop the overall estimate.

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

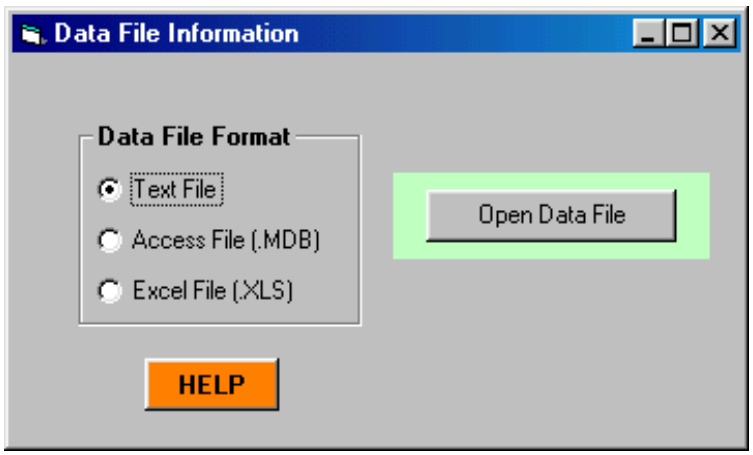
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

Universe size

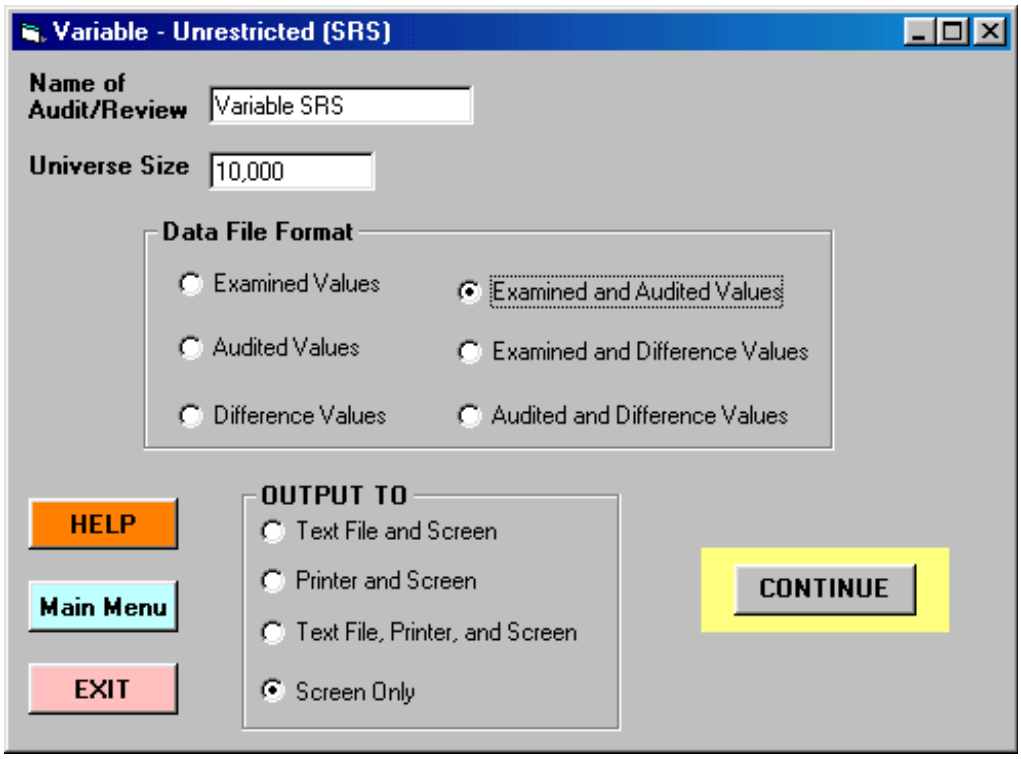
The universe size is the total number of items from which the sampled items were selected. The number should be entered without commas but upon exiting this box, the commas will be inserted. This number will be used in estimating universe parameters.

Specify input file

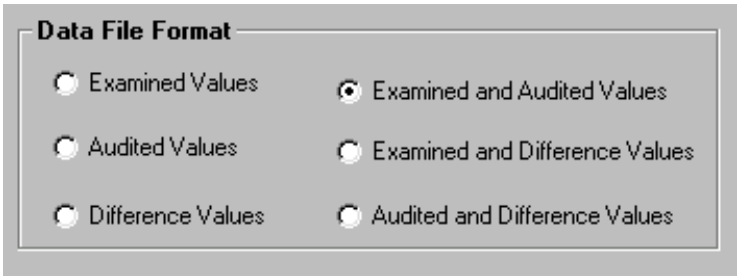
After entering the above information, click on **Specify Input File**. The following screen will appear. The input file format can be a text file, a table within an Access database, or an Excel spreadsheet. After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open"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



When the data file has been opened, the full input screen will appear (shown next).



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. 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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Input From a Text File

The sample data 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).

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

Data set DATASRS.TXT

```
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
 .
 .
 .
48   300  237
49   500  435
50   100  86
```

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

When the data file has been opened, the program will return to the input screen.

Input From an Access Database

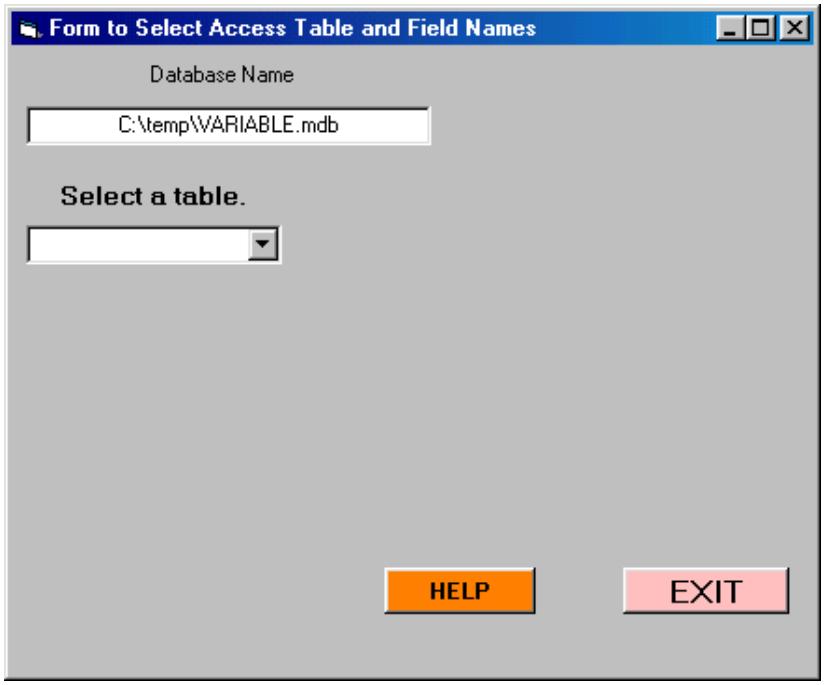
The sample data may be stored in a table within an Access database. Select the name of the database containing the input table in the preceding **Open Data File** step. This database must have the standard Access extension (.MDB). The name of the database for this illustration is C:\TEMP\VARIABLE.MDB. The 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.

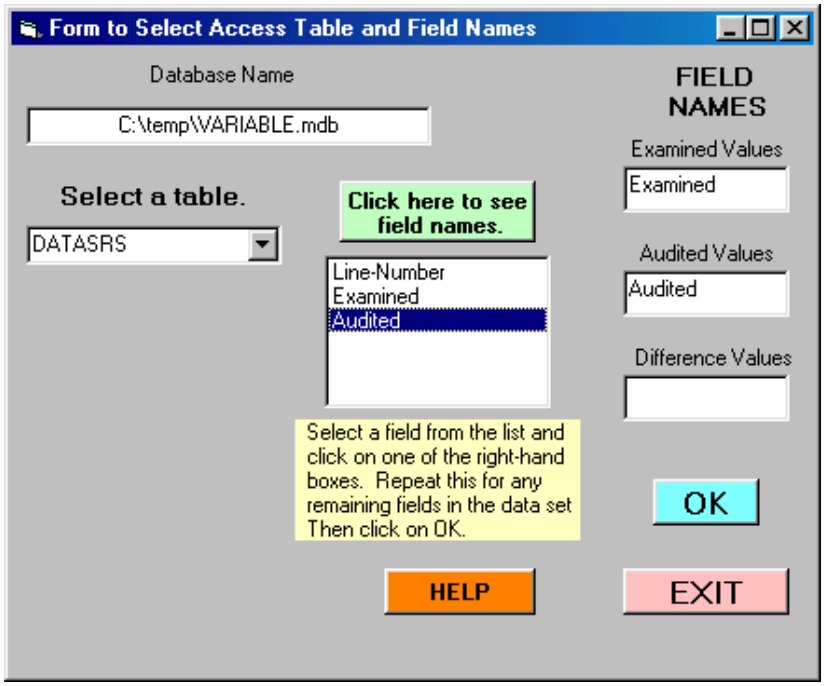
DATASRS : Table

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 to the organization's minimum sample size standards.

After opening the database, 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 return to the input screen. **NOTE**: When the user returns to the 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 input screen.

Input From an Excel Spreadsheet

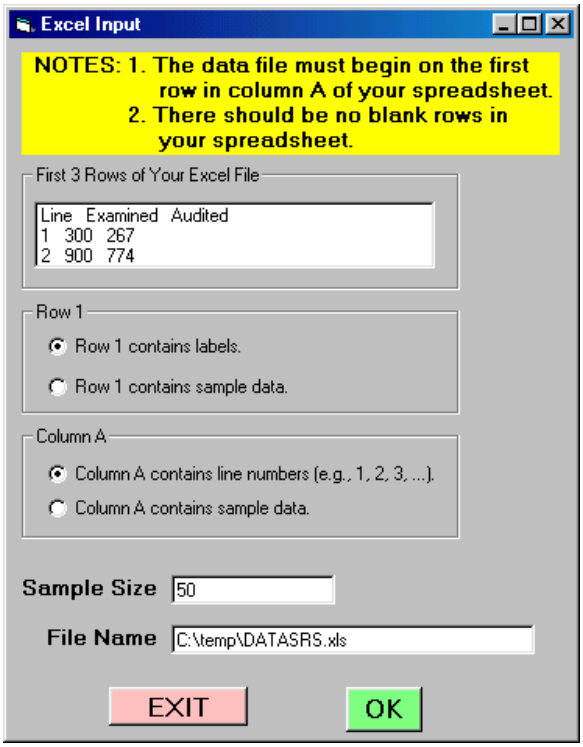
The sample data can be stored in an Excel spreadsheet. Select the name of the spreadsheet containing the input table in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATASRS.XLS will be used. The first 20 rows of this file are shown next.

	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

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

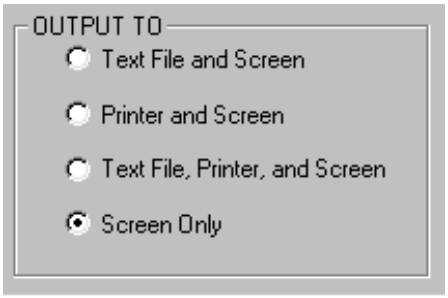
After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will appear. This particular file contains labels (variable names) in the first row and line numbers (1, 2, 3, ...) in

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.



When the Excel Input information has been completed, click on **OK**. The program will resume processing.

Output Options

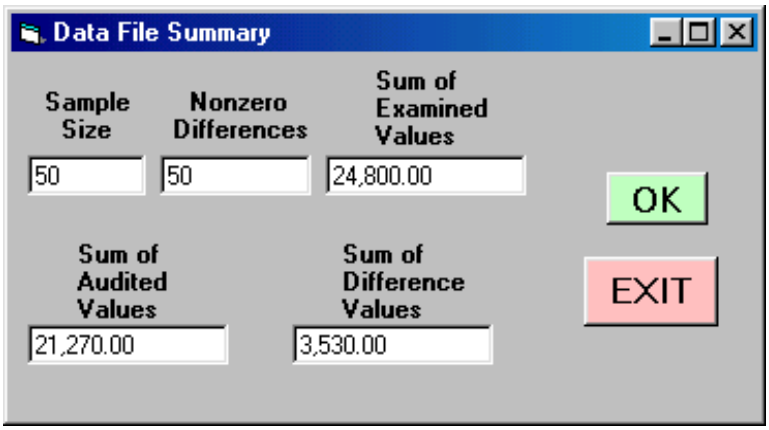


The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

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.

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

For the examined, adjusted, and difference sections of the output, the following pieces of information will be displayed. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

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	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 (MEAN)	A measurement of the variation of the estimated universe mean with respect to all possible estimated means for this universe and sample size.

STANDARD ERROR (TOTAL)	A measurement of the variation of the estimated universe total with respect to all possible estimated totals 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	The result of dividing the precision amount by the point estimate and stating the result as a percentage.
t-VALUE USED	The t- percentile value used to construct the confidence interval.

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.

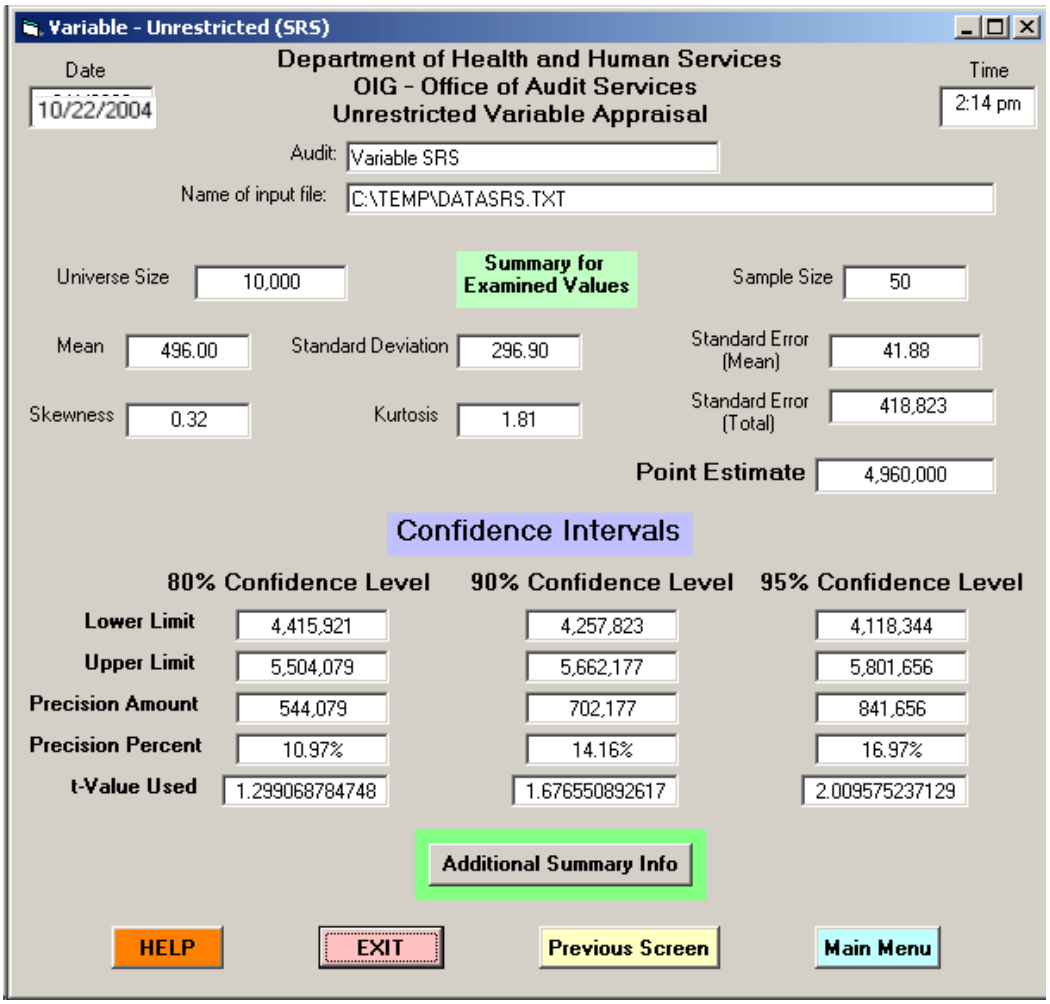
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004         VARIABLE UNRESTRICTED APPRAISAL          Time: 14:14
                          AUDIT/REVIEW: Variable SRS
                     DATA FILE USED: C:\temp\DATASRS.TXT

     SAMPLE         EXAMINED    NONZERO        TOTAL OF          TOTAL OF
      SIZE            VALUE      DIFFS        DIFF VALUES        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
          SKEWNESS                                    .32
          KURTOSIS                                   1.81
          STANDARD ERROR (MEAN)                     41.88
          STANDARD ERROR (TOTAL)                  418,823
          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
          SKEWNESS                                    .30
          KURTOSIS                                   1.78
          STANDARD ERROR (MEAN)                     36.14
          STANDARD ERROR (TOTAL)                  361,412
          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
SKEWNESS                                    .64
KURTOSIS                                   2.98
STANDARD ERROR (MEAN)                      6.81
STANDARD ERROR (TOTAL)                   68,068
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
```

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

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**.



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

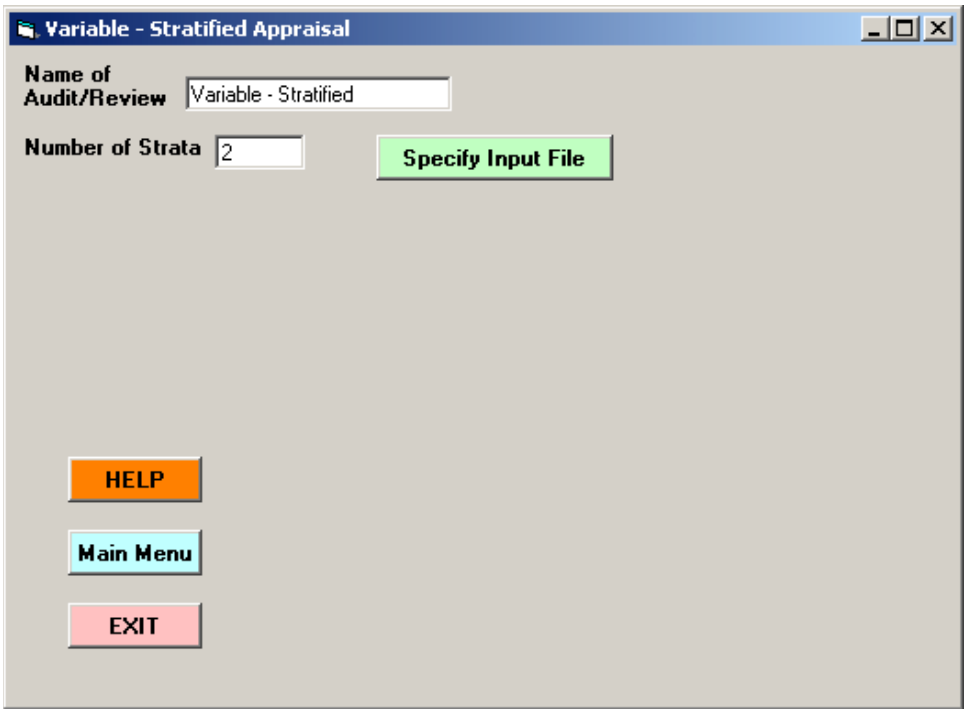
STRATIFIED

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

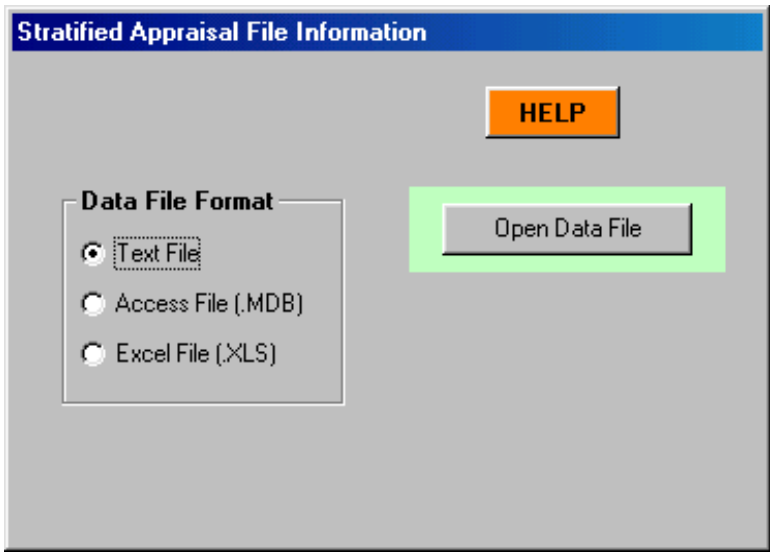
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

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.

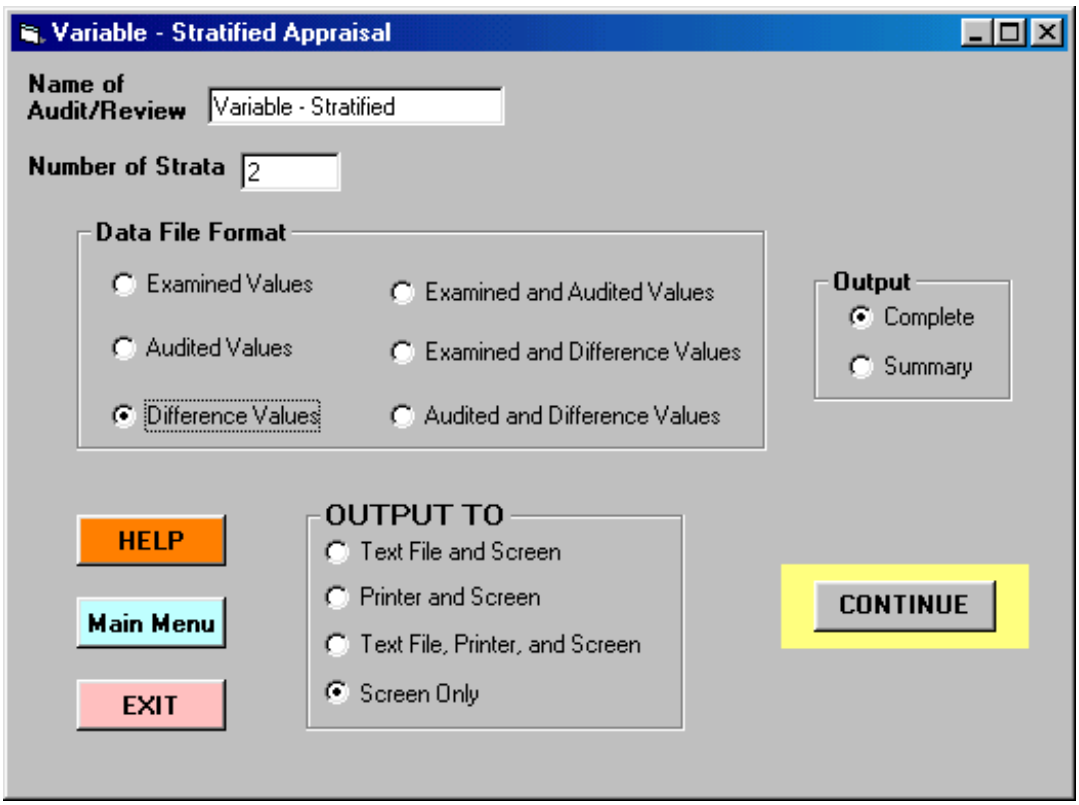
Specify input file

After entering the name and the number of strata, click on the **Specify Input File** button and the following screen will appear. The input file(s) format can consist of two text files, two tables within an Access database, or two spreadsheets within the same Excel file.

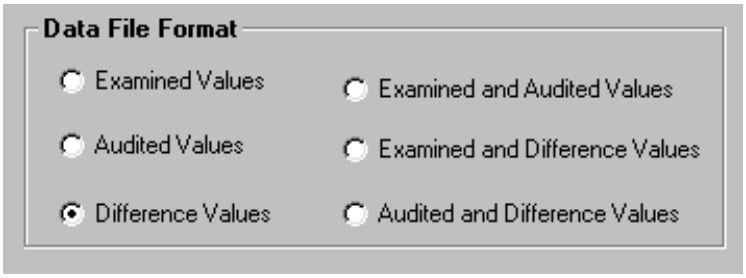


After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open" file screen, locate the input file and double-click on it.

When all files have been opened, the full input screen (shown next) will appear:



Format of Input File



Prior to executing this program, the user must create a sample 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 sample 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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Input From a Text File

The sample data and universe/sample size 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).

Sample Data File

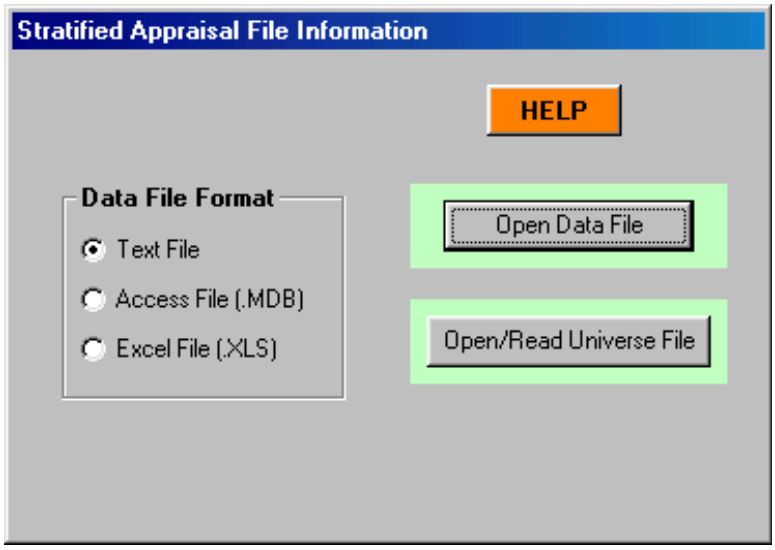
A portion of data set DATASTRAT.TXT is shown below. The sample data file contains 25 observations in each stratum.

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

Sample Data file DATASTRAT.TXT

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

After clicking on the **Open Data File** button and opening the sample data file (e.g., DATASTRAT.TXT), the **Open/Read Universe File** button will appear (shown next).

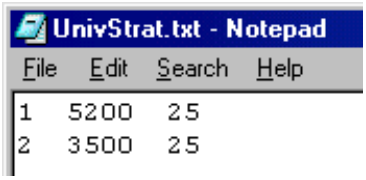


Universe/Sample Size File

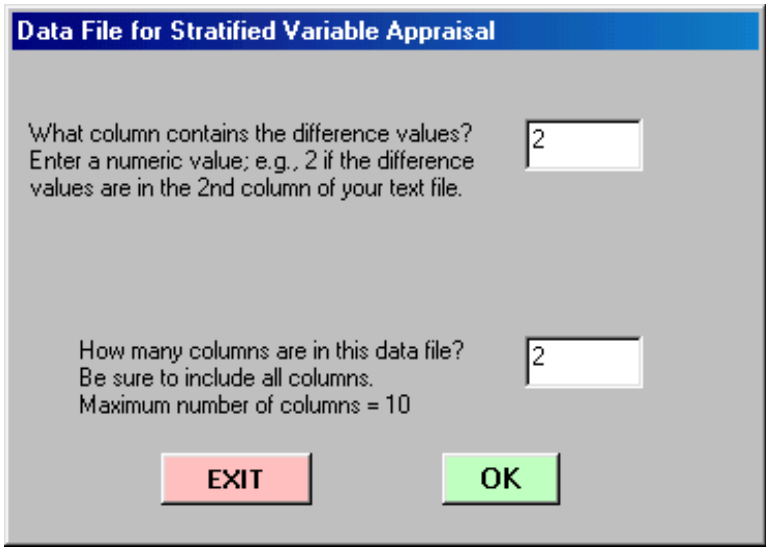
To open the file containing the universe/sample sizes, click on the **Open/Read Universe File** button and use the standard Windows "Open" file screen to locate the universe file (e.g., UnivStrat.TXT--shown below). The format of this file is:

line counter, universe size, sample size

There is one line for each stratum in the sample. Values within a line can be separated by one or more spaces or by using the tab key. Commas are allowed in the universe and sample size values. After the universe file has been selected, the program will return to the input screen.



After opening the data files, the user is returned to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input sample data and the total number of columns in the sample data file using the screen below:



After entering the column information, click on **OK**. The program will resume processing.

Input From an Access Database

The sample data and universe/sample size information must be stored in two tables within the same Access database. Select the name of the database containing the input tables 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 on the standard Windows "Open" file screen.

Universe/Sample Sizes Table

The name of the Access table containing the universe/sample size information for this illustration is UnivStrat. The following table shows the table contents:

UnivStrat : Table

Counter	Universe Size	Sample Size
1	5200	25
2	3500	25

Sample Data Table

The name of the Access table containing the sample data 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.

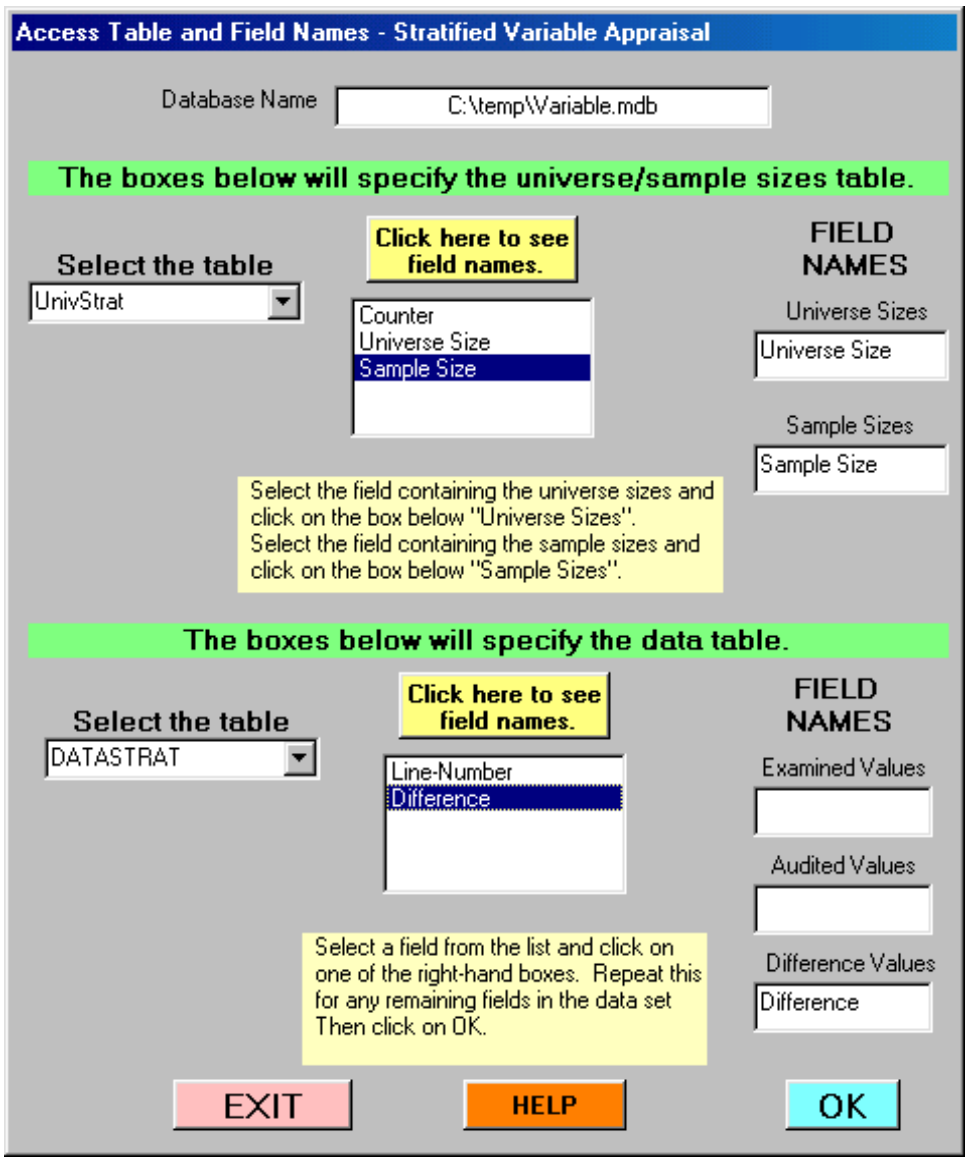
DATASTRAT : Table

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

This sample data file contains 25 observation in each stratum. The last three rows of the first stratum and the first two rows of the second stratum are shown below:

23	102
24	69
25	76
26	354
27	328

After opening the database, the user will be asked to select the name of the table containing the universe/sample information and the table containing the sample data using the following form. For the universe/sample sizes table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (UnivStrat for this illustration), click on **Click here to see field names**. This list will contain the field names for this table. To select the field names, click on the field name for the field containing the universe sizes ("Universe Size" in this illustration) and click on the box labeled "Universe Sizes." The field name will appear in this box. Repeat this procedure for the field containing the sample sizes ("Sample Size" in this illustration).



To select the field names for the table containing the sample data, first select the table (DATASTRAT in this illustration), click on **Click here to see field names**, and 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 return to the input screen. **NOTE**: When the user returns to the input screen, the data file format (Difference Values for this illustration) will be selected, based on responses within the preceding

Access Table and Field Names window. When using an Access input file, the user cannot change this data file format option after returning to the input screen.

Input From an Excel Spreadsheet

With this option, the sample data and universe/sample sizes information must be stored in two Excel spreadsheets within the same Excel file. Select the name of the file containing both spreadsheets 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.

Universe/Sample Sizes Spreadsheet

The contents of the spreadsheet containing the universe/sample sizes (named Universe in this illustration) are shown below:

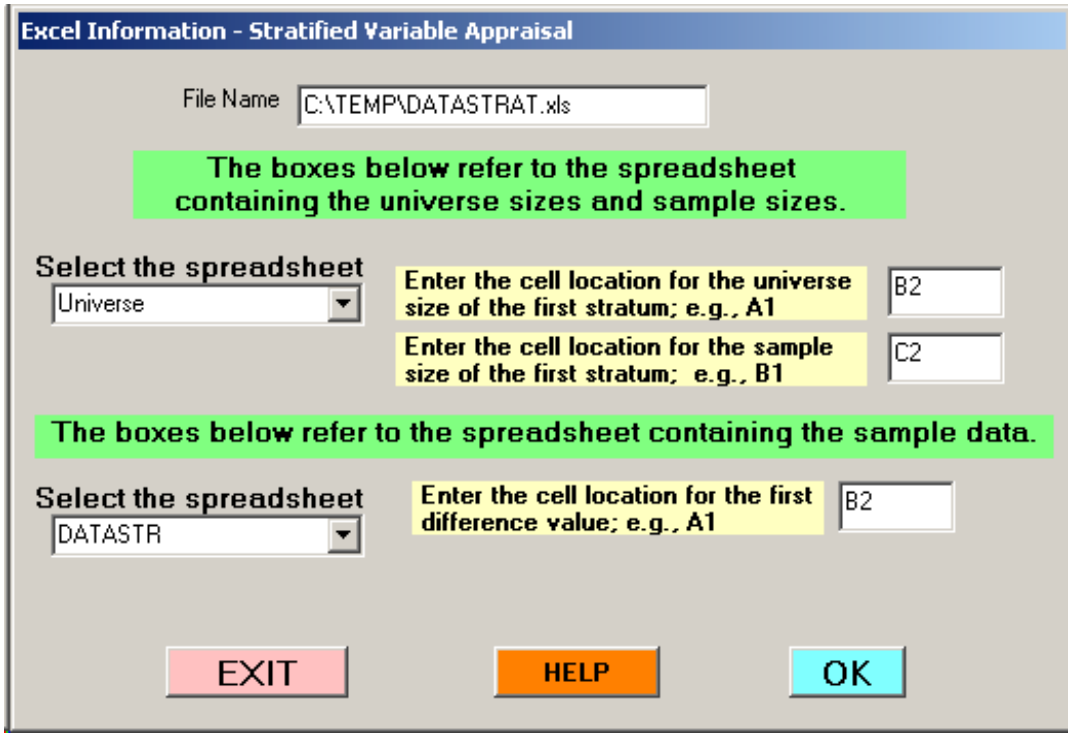
	A	B	C
1	Stratum	Universe Size	Sample Size
2	1	5200	25
3	2	3500	25

Sample Data Spreadsheet

The name of the second spreadsheet containing the sample data is DATASTR in this illustration. The first five observations in this spreadsheet are shown below. The sample data file contains 25 observations in each stratum for a total of 51 rows (including the first row with labels). This particular file contains line numbers (1, 2, 3, . . .) in column A. The line numbers are optional.

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

After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will appear. For this illustration, the various boxes should be filled in as shown.



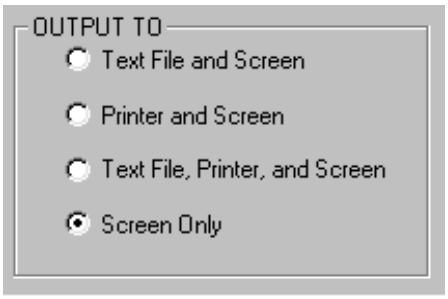
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

After entering the cell locations, click on **OK**. The program will resume processing.

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

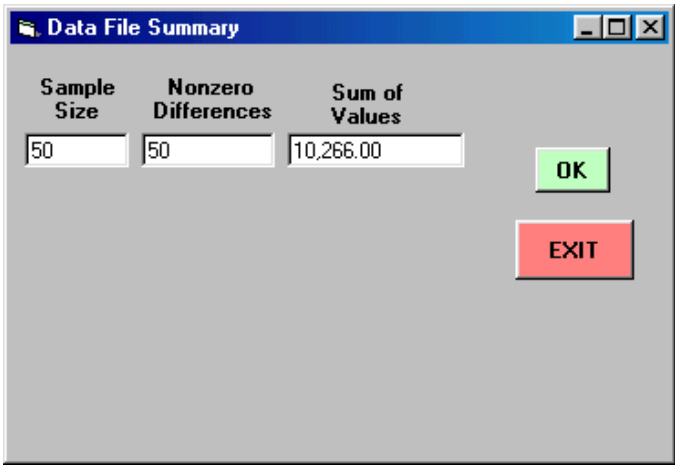


The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

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.

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 sample data file is complete and accurate. Click on **OK** to continue or **EXIT** to exit the program.



Program Output

For the examined, audited, and difference sections of the output, the following pieces of information will be displayed. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

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	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 (MEAN)	A measurement of the variation of the sample mean with respect to all possible means for this stratum universe and this sample size.
STRATUM STANDARD ERROR (TOTAL)	A measurement of the variation of the estimated stratum universe total with respect to all possible estimated totals for this stratum universe and 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	The result of dividing the precision amount by the point estimate and stating the result as a percentage.
t-VALUE USED	The t-percentile value used to construct the confidence interval.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

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.

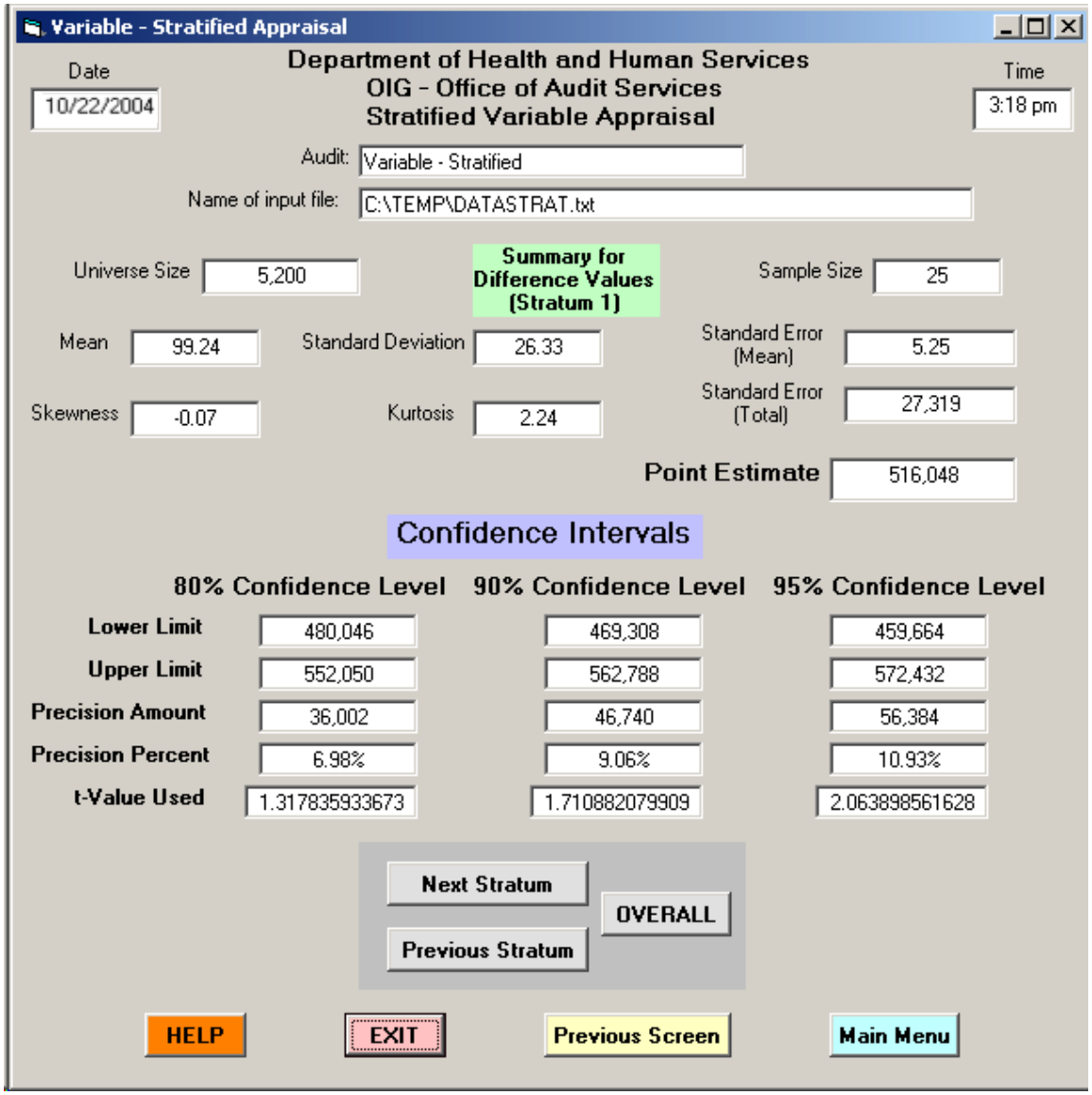
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004          STRATIFIED VARIABLE APPRAISAL           Time: 15:18
                     AUDIT/REVIEW: Variable - Stratified

                    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
           SKEWNESS                                  -.07
           KURTOSIS                                  2.24
           STANDARD ERROR (MEAN)                     5.25
           STANDARD ERROR (TOTAL)                  27,319
           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
           SKEWNESS                                  -.06
           KURTOSIS                                  1.85
           STANDARD ERROR (MEAN)                     7.90
           STANDARD ERROR (TOTAL)                  27,651
           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
```

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

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 sample 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 since only the difference values were contained in the sample 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**.



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

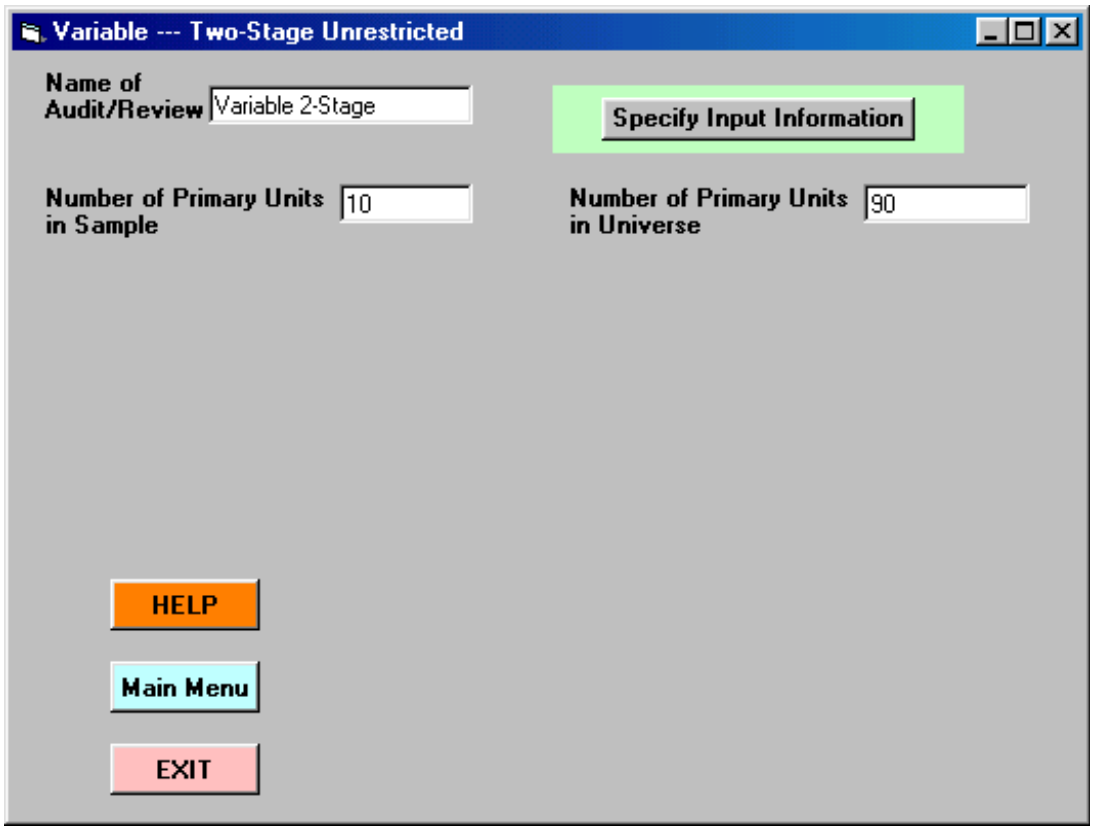
TWO-STAGE UNRESTRICTED

Purpose

This program performs a two-stage variable appraisal on a data file previously created by the user based on information gathered from a two-stage 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 the appraisal program.

Input Screen

NOTE: The **Specify Input Information** button only becomes visible when a value has been specified in the **Number of Primary Units in Universe** box.



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.

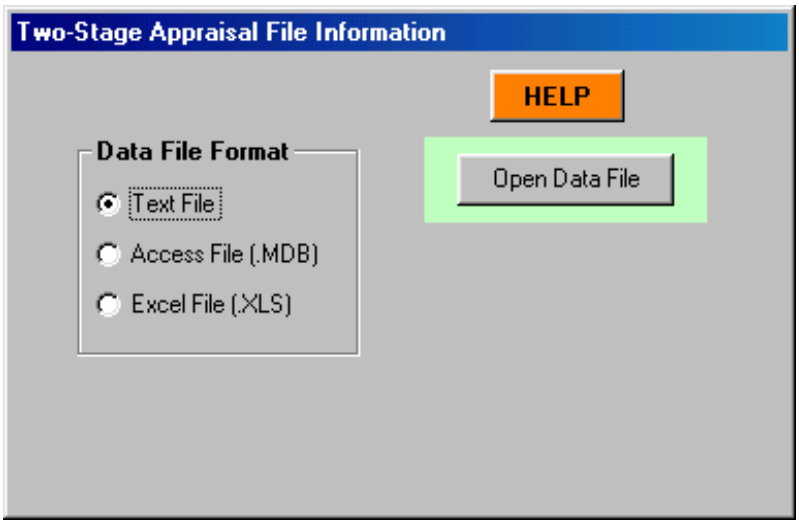
Number of primary units in sample

The user enters the number of primary units in the sample. The program will only accept a numeric entry.

Number of primary units in universe

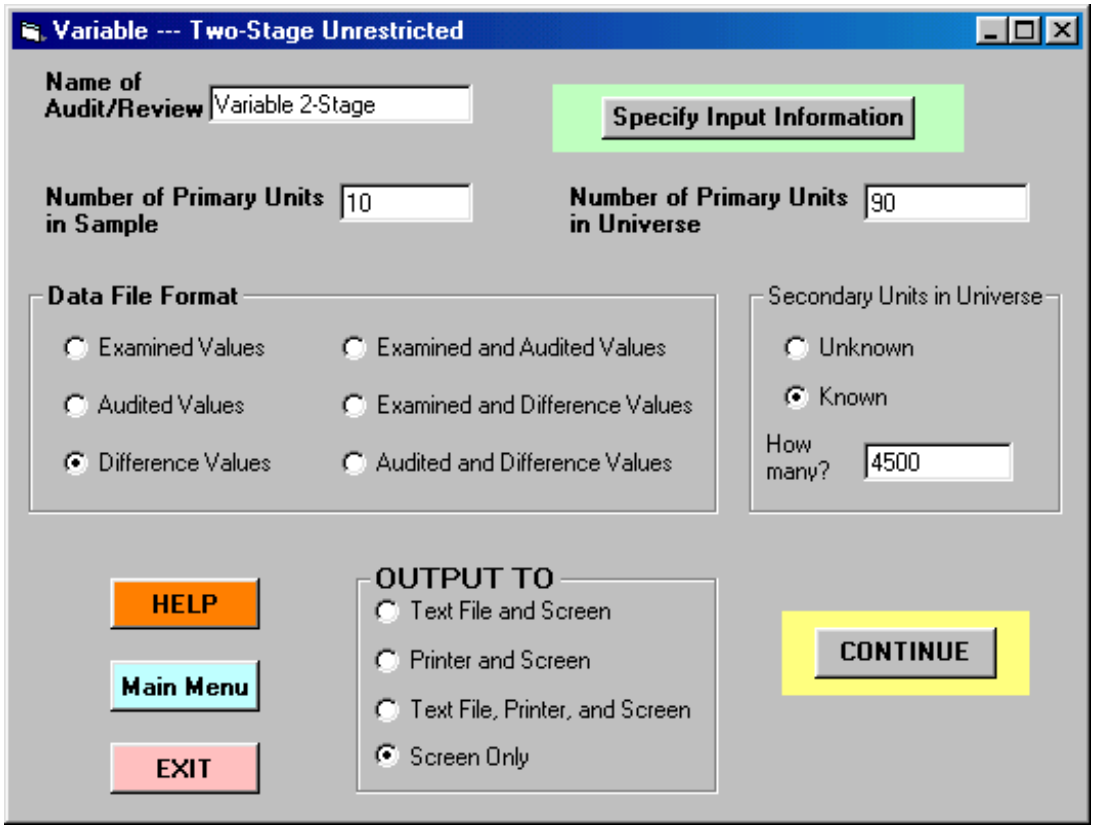
The user enters the number of primary units in the universe. The program will only accept a numeric entry.

After entering the above information, click on **Specify Input Information**. The following screen will appear. The input file(s) format can consist of two text files, two tables within an Access database, or two spreadsheets within the same Excel file. If the text file option is selected, a series of buttons will become visible. To specify the data file click on **Open Data File** and to open the file containing the universe/sample sizes click on the **Open/Read Universe File** button. If either the Access File or Excel File option is selected, click on **Open Data File** to specify the Access database or Excel file.

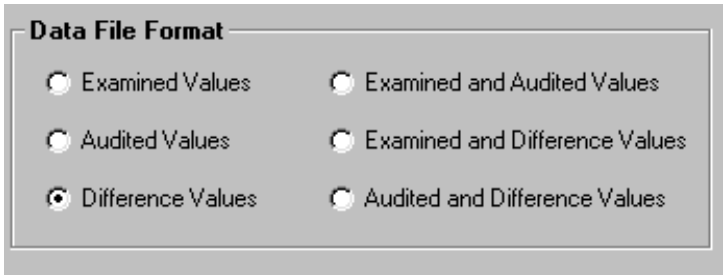


After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open" file screen, locate the input file and double click on it.

When all files have been opened, the full input screen will appear.



Format of Input File



Prior to executing this program, the user must create a sample 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 sample 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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Input From Text Files

The sample data and universe/sample size 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).

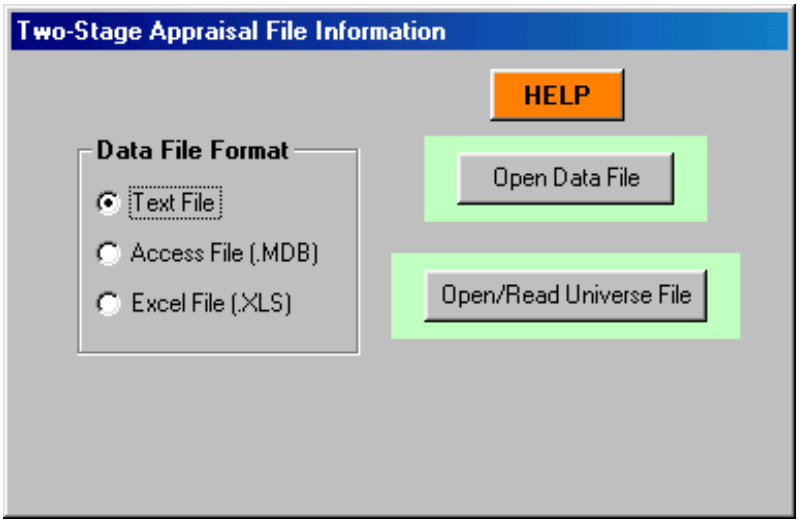
Sample Data File

The sample data file used in this illustration (C:\TEMP\DATA2STG.TXT) is shown below and consists of data from 10 sampled primary units selected from a total of 90 primary units. There are a total of 104 sample items.

Page 1		*Page 2*		*Page 3*	*Page 4*		*Page 5*	
1	5	22	1	4311	64	4	85	3
2	7	23	5	444	65	3	86	6
3	9	24	5	453	66	7	87	4
4	0	25	6	461	67	6	88	7
5	11	26	4	470	68	7	89	3
6	2	27	11	482	69	8	90	9
7	8	28	12	498	70	4	91	1
8	4	29	0	506	71	3	92	4
9	3	30	1	515	72	2	93	5
10	5	31	8	523	73	3	94	6
11	4	32	4	5312	74	6	95	7
12	3	33	6	5411	75	4	96	5
13	7	34	4	553	76	3	97	10
14	2	35	0	564	77	2	98	11
15	11	36	1	572	78	2	99	2
16	0	37	0	580	79	8	100	1
17	1	38	9	590	80	4	101	4
18	9	39	8	601	81	0	102	0
19	4	40	4	614	82	4	103	5
20	3	41	6	623	83	5	104	4
21	2	42	10	632	84	6		

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

After clicking on the **Open Data File** button and opening the sample data file (e.g., DATA2STG.TXT), the **Open/Read Universe File** button will appear on the following screen.

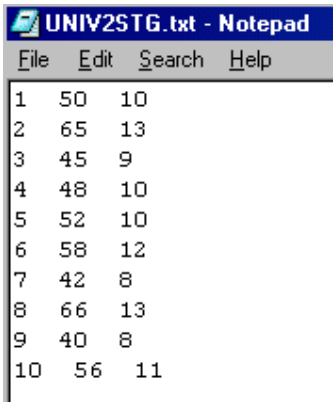


Universe/Sample Size File

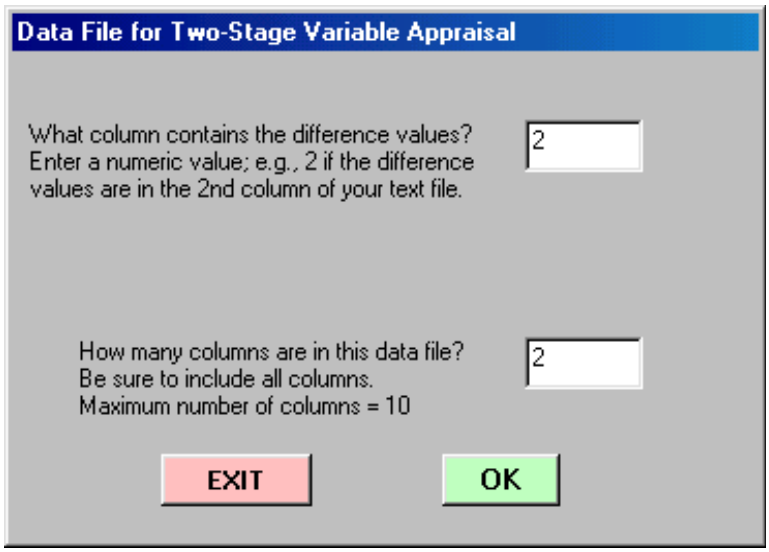
To open the file containing the universe/sample sizes, click on the **Open/Read Universe File** button and use the standard Windows "Open" file screen to locate the universe file (e.g., UNIV2STG.TXT, shown below). The format of this file is:

line counter, universe size, sample size

There is one line for each primary unit in the sample. Values within a line can be separated by one or more spaces or by using the tab key. Commas are allowed in the universe and sample size values.



After selecting the universe file and clicking on **OK**, the program will return to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input data and the total number of columns in the sample data file using the screen below:



After entering the column information, click on **OK**. The program will resume processing.

Input From an Access Database

The sample data and universe/sample size information must be stored in two tables within the same Access database. Select the name of the database containing the input tables 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.

Universe/Sample Size Table

The name of the Access table containing the universe/sample size information for this illustration is Univ2Stg. The following table shows the table contents.

Univ2Stg : Table

Counter	Universe Size	Sample Size
1	50	10
2	65	13
3	45	9
4	48	10
5	52	10
6	58	12
7	42	8
8	66	13
9	40	8
10	56	11

Sample Data Table

The first 14 rows of the Access table containing the sample data DATA2STG are shown next. The field name for the first column ("Line" in the illustration) is arbitrary and is not specified or used by the program at any point. For a view of the complete file, refer to the previous **Input From a Text File** section.

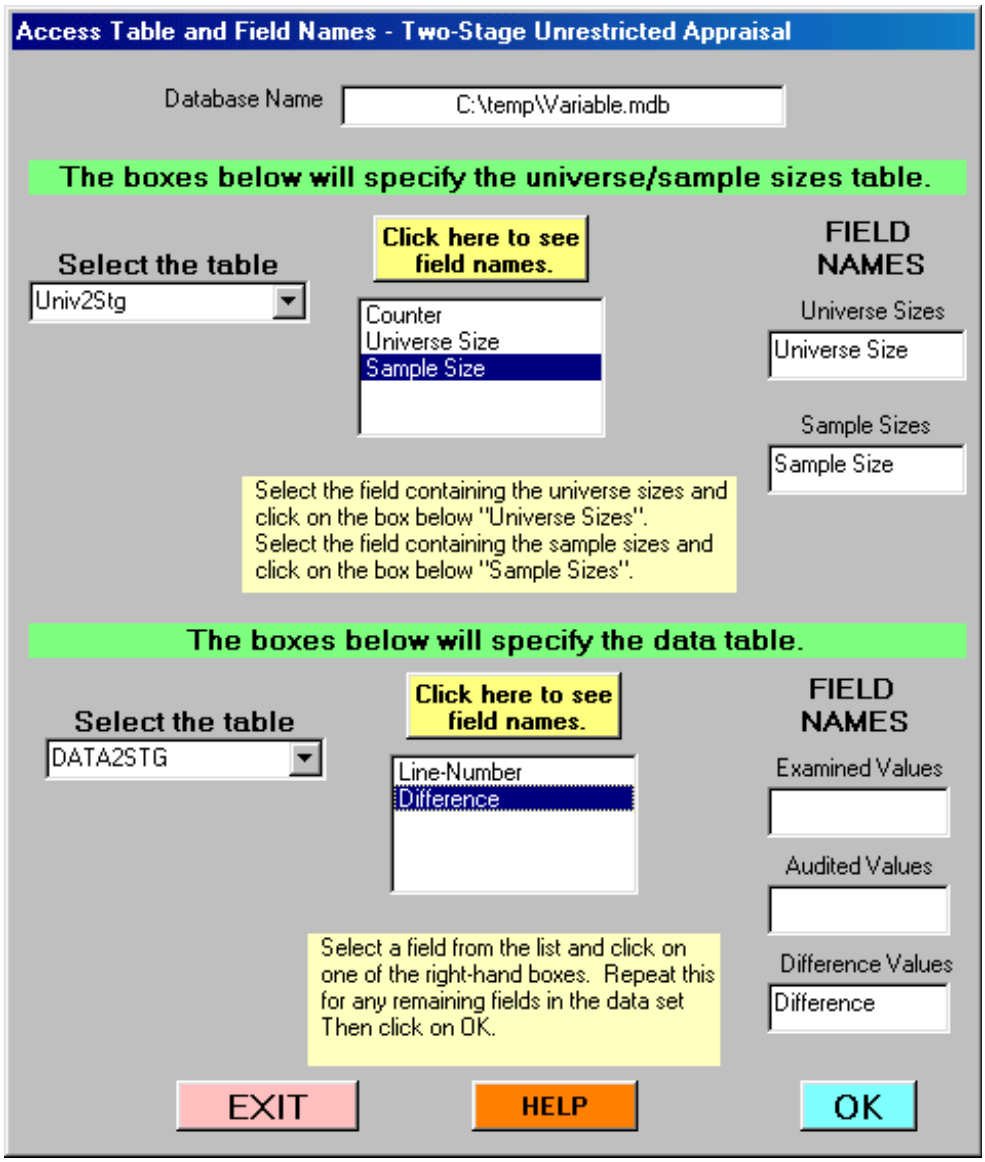
DATA2STG : Table

Line	Difference
1	5
2	7
3	9
4	0
5	11
6	2
7	8
8	4
9	3
10	5
11	4
12	3
13	7
14	2

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

After opening the database, the user will be asked to select the name of the table containing the universe/sample information and the table containing the sample data using the following form. For the universe/sample sizes table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (Univ2Stg for this illustration), click on **Click here to**

see field names. This list will contain the field names for this table. To select the field names, click on the field name for the field containing the universe sizes ("Universe Size" in this illustration) and click on the box labeled "Universe Sizes." The field name will appear in this box. Repeat this procedure for the field containing the sample sizes ("Sample Size" in this illustration).



To select the field name(s) for the sample data, first select the table (DATA2STG in this illustration), click on **Click here to see field names**, and click on the field name for the field containing the first piece of information in the input file ("Difference" in this illustration) and then click on the box labeled "Difference." The field name will then 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 return to the input screen. **NOTE**: When the user returns to the 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 input screen.

Input From Excel Spreadsheets

With this option, the sample data file and universe/sample sizes information must be stored in two Excel spreadsheets within the same Excel file. Select the name of the file containing both spreadsheets in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS).

Universe/Sample Size Spreadsheet

For this illustration, Excel file DATA2STG.XLS will be used. The contents of the spreadsheet containing the universe/sample sizes (named Universe in this illustration) are shown below:

	A	B	C
1	Counter	Universe	Sample
2	1	50	10
3	2	65	13
4	3	45	9
5	4	48	10
6	5	52	10
7	6	58	12
8	7	42	8
9	8	66	13
10	9	40	8
11	10	56	11

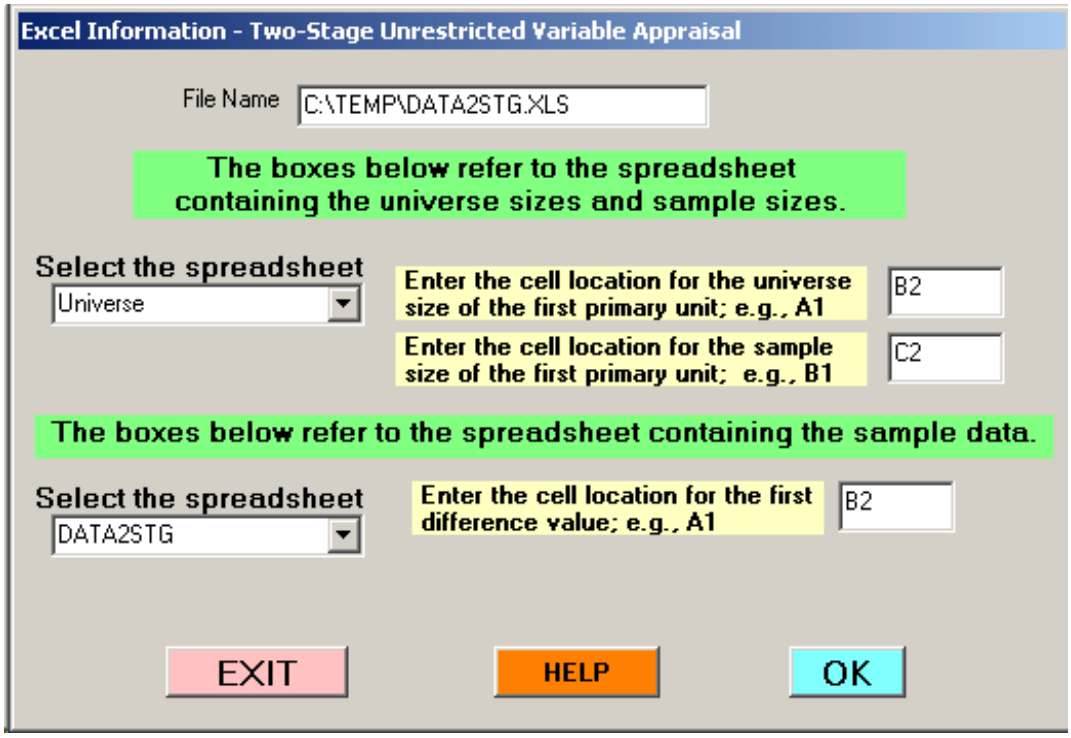
Sample Data Spreadsheet

The name of the second spreadsheet containing the sample data is DATA2STG in this illustration. The first 15 observations in this file are shown next. For a view of the complete file, refer to the previous **Input From a Text File** section. This particular file contains line numbers (1, 2, 3, . . .) in column A. The line numbers are optional.

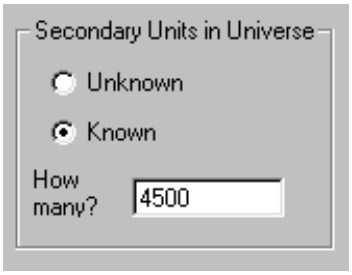
	A	B
1	Line	Difference
2	1	5
3	2	7
4	3	9
5	4	0
6	5	11
7	6	2
8	7	8
9	8	4
10	9	3
11	10	5
12	11	4
13	12	3
14	13	7
15	14	2
16	15	11

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

After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will be displayed. For this illustration, the various boxes should be filled in as shown. After entering the cell locations, click on **OK**. The program will resume processing.

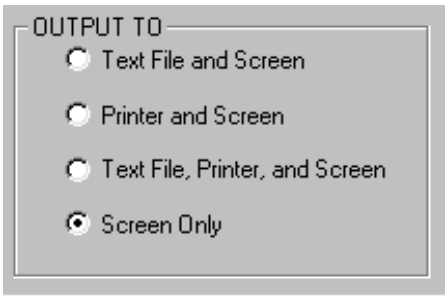


Specifying the Number of Secondary Units



The user enters the total number of secondary units for all of the primary units in the universe. The program will only accept a numeric entry. If the number of secondary units is not known, click on **Unknown**. For this case, the output will not display the quantity of secondary units in the universe.

Output Options

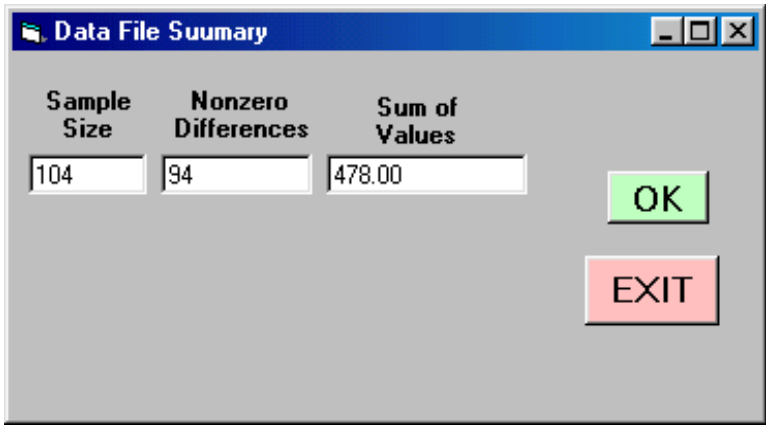


The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

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.

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

For the examined, adjusted, and difference sections of the output, the following pieces of information will be displayed. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

UNIT NBR	The indicator of the sampled primary unit that is being appraised.
SAMPLE SIZE	The number of items sampled in the particular primary unit.
NONZERO ITEMS	The number of nonzero items sampled in the particular primary unit.
SAMPLE 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.
VARIANCE	A measurement of the variation of the sampled items about the sample mean.
UNIVERSE SIZE	The quantity of the secondary items within each sampled primary unit from which the sample was drawn. The results of the sample will be projected to the primary unit universe using this value.
POINT ESTIMATE	The single estimate for the universe total of the primary unit based on the sample mean and universe size.
STANDARD ERROR	A measurement of the variation of the overall point estimate of the universe total with respect to all possible point estimates for this universe and these sample sizes.
CONFIDENCE LEVELS	The confidence (80%, 90%, or 95%) associated with the ability of 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 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 The result of dividing the precision amount by the point estimate.

Z-VALUE USED The value referred to as the standard normal deviate. It is a measurement from the point estimate to a confidence limit measured in standard errors.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUT2STG.TXT, shown next. The printer output is identical.

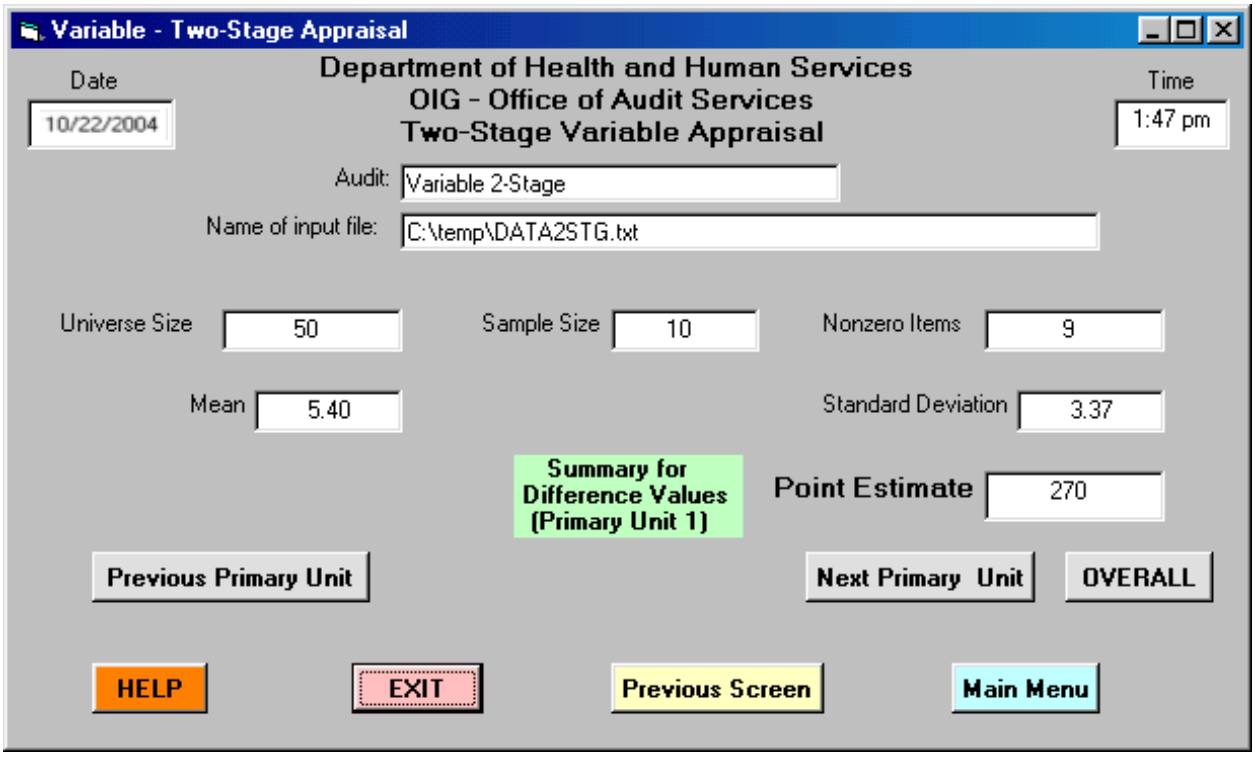
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                         OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004   TWO-STAGE UNRESTRICTED VARIABLE APPRAISAL       Time: 13:47
                         AUDIT/REVIEW: Variable 2-Stage

                      DATA FILE USED: C:\temp\DATA2STG.TXT

           --------------------- D I F F E R E N C E ----------------------
UNIT   SAMPLE SIZE/
 NBR   NONZERO ITEMS   SAMPLE MEAN     VARIANCE   UNIVERSE SIZE  POINT ESTIMATE

  1        10/9              5.40        11.38             50             270
  2        13/12             4.00        10.67             65             260
  3         9/8              5.67        16.75             45             255
  4        10/8              4.80        13.29             48             230
  5        10/9              4.30        11.12             52             224
  6        12/10             3.83        14.88             58             222
  7         8/8              5.00         5.14             42             210
  8        13/12             3.85         4.31             66             254
  9         8/8              4.88         6.13             40             195
 10        11/10             5.00        11.80             56             280

          104/94             4.80                          522           2,400
```

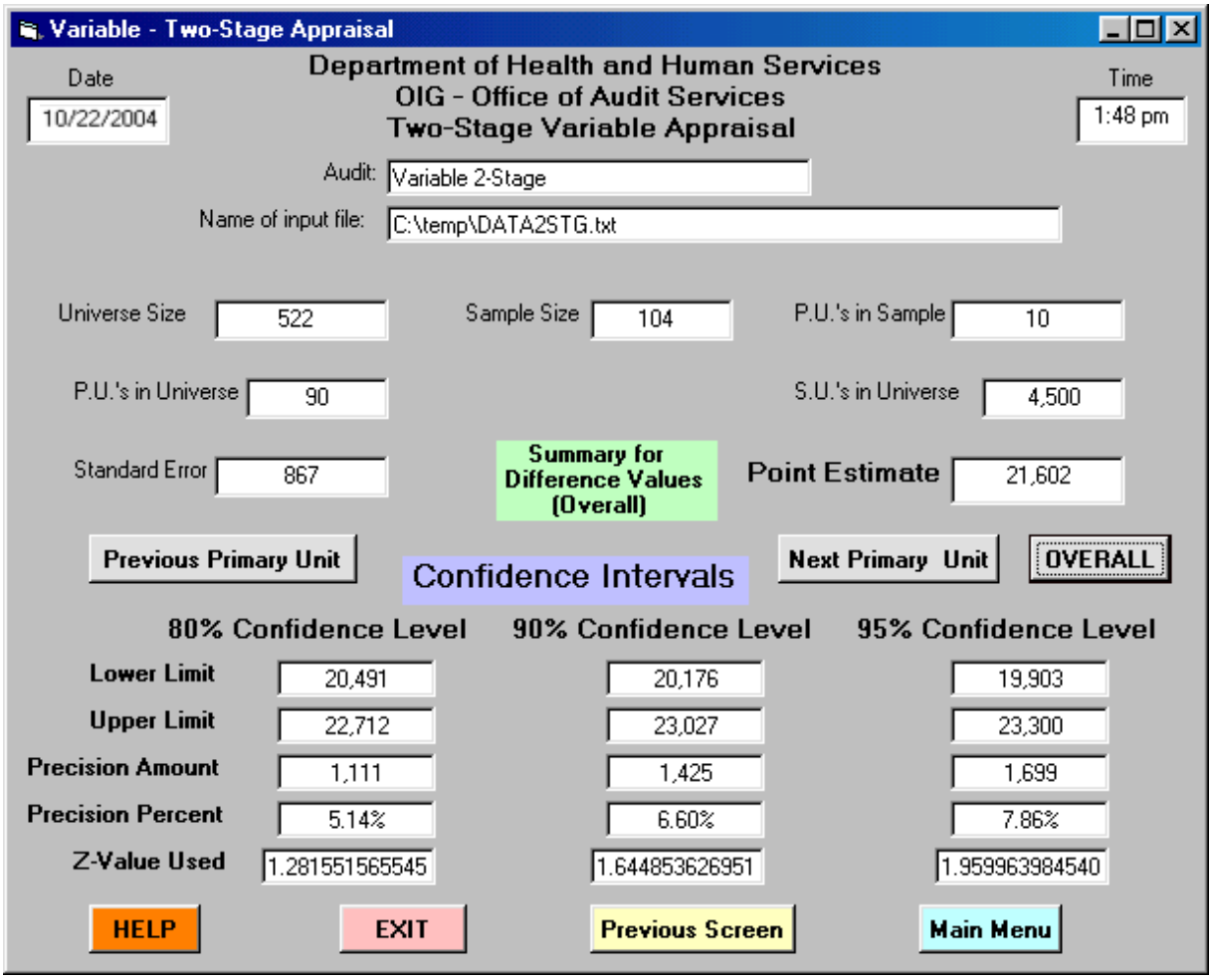
```
NOT SAMPLED
    80                                                  3,978
OVERALL TOTALS
    90                                                  4,500          21,602

STANDARD ERROR                                                            867

                                        CONFIDENCE LIMITS
                                    80% CONFIDENCE LEVEL
           LOWER LIMIT                                 20,491
           UPPER LIMIT                                 22,712
           PRECISION AMOUNT                             1,111
           PRECISION PERCENT                             5.14%
           Z-VALUE USED                        1.281551565545

                                    90% CONFIDENCE LEVEL
           LOWER LIMIT                                 20,176
           UPPER LIMIT                                 23,027
           PRECISION AMOUNT                             1,425
           PRECISION PERCENT                             6.60%
           Z-VALUE USED                        1.644853626951

                                    95% CONFIDENCE LEVEL
           LOWER LIMIT                                 19,903
           UPPER LIMIT                                 23,300
           PRECISION AMOUNT                             1,699
           PRECISION PERCENT                             7.86%
           Z-VALUE USED                        1.959963984540
```

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

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 screen below is the summary of the difference values for the first primary unit in this illustration. If the user created a sample 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 form. For this illustration, the **Additional Summary Info** button is not visible since only the difference values were used in the data file. To obtain the results for the second primary unit, click on **Next Primary Unit**. The user can click on **Next Primary Unit** and **Previous Primary Unit** to review the results for the individual primary units. To obtain the overall results, click on **OVERALL** (shown next, immediately following the summary screen for the first primary unit).


Variable - Two-Stage Appraisal
Date
10/22/2004
Department of Health and Human Services
OIG - Office of Audit Services
Two-Stage Variable Appraisal
Time
1:47 pm
Audit:
Variable 2-Stage
Name of input file:
C:\temp\DATA2STG.txt
Universe Size
50
Sample Size
10
Nonzero Items
9
Mean
5.40
Standard Deviation
3.37
Summary for Difference Values (Primary Unit 1)
Point Estimate
270
Previous Primary Unit
Next Primary Unit
OVERALL
HELP
EXIT
Previous Screen
Main Menu



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

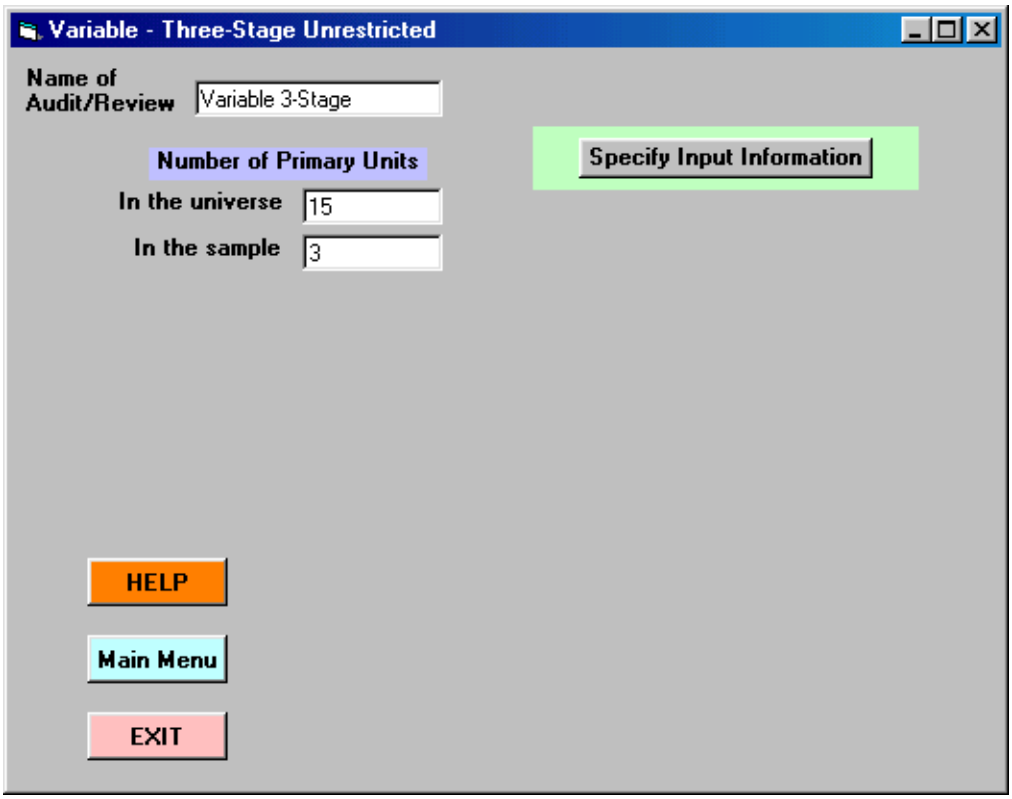
THREE-STAGE UNRESTRICTED

Purpose

This program performs a three-stage variable appraisal on a data file previously created by the user based on information gathered from a three-stage 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 the appraisal program.

Input Screen

NOTE: The **Specify Input Information** button only becomes visible when the number of primary units in the universe and the sample have been entered.



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.

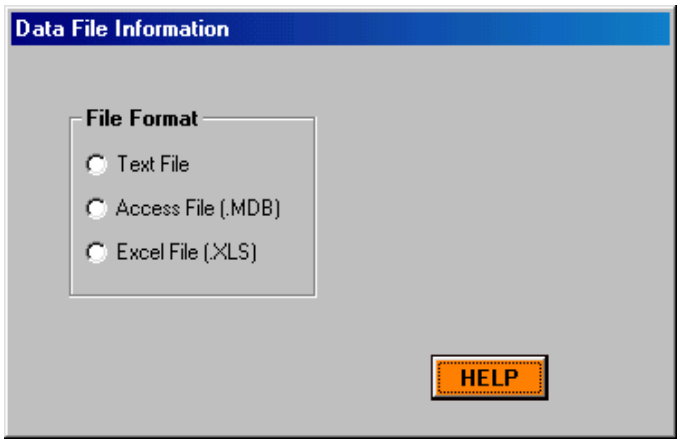
Number of primary units in the universe

The user enters the number of primary units in the universe. This quantity may be entered with commas and the program will only accept a numeric entry.

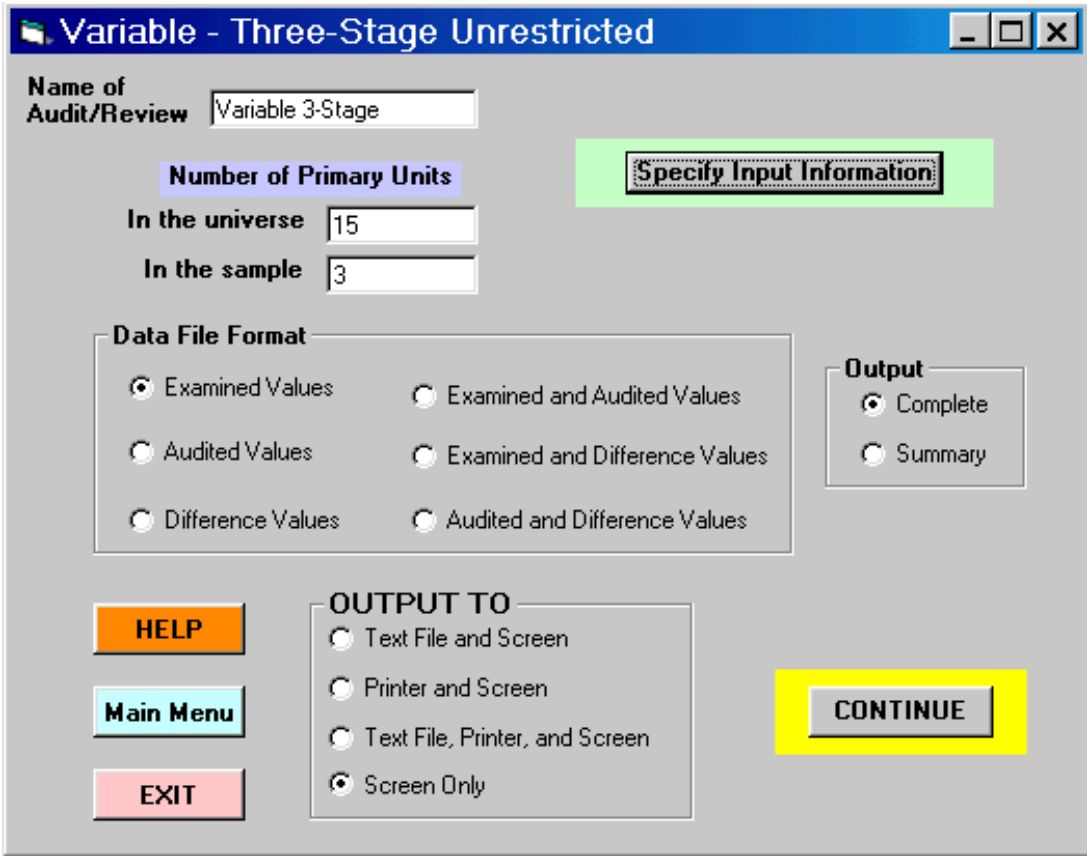
Number of primary units in the sample

The user enters the number of primary units in the sample. This quantity may be entered with commas; the program will only accept a numeric entry.

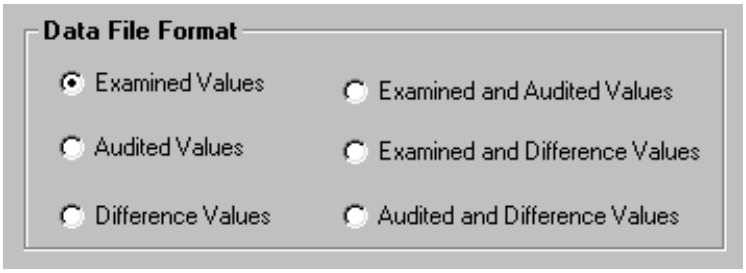
After entering the above information, click on **Specify Input Information**. The following screen will appear. The format can consist of three text files, three tables within the same Access database, or three spreadsheets within the same Excel file. If the text file option is selected, a series of buttons will become visible (one at a time) and the user will need to click on **Open Data File** to specify the data file, click on the **Open Primary Unit File** button to specify the file containing information on the primary units, then click on the **Open Secondary Unit File** button to specify the file containing information on the secondary units. If either the Access File or Excel File option is selected, click on **Open File** to specify the Access database or Excel file.



When all files have been opened, the full input screen will appear.



Format of Input File



Prior to executing this program, the user must create a sample 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.

Format of Primary Unit Information

Regardless of the software used to create the primary unit file (a text file, Access table, or Excel spreadsheet), each line in the this file must contain at least three variables (fields). One possible format is the following:

REGION 1 10 2

Explanation:

REGION 1 - This is a description of a primary unit that was sampled. The description may be up to 25 characters in length and may include spaces, commas, and other special characters.

10 - This is the universe size of secondary units within this primary unit.

2 - This is the number of secondary units sampled from this primary unit.

Primary Unit File Restrictions

1. The primary unit description, the universe size of secondary units, and the number of sampled secondary units must be the final three entries in each line.

2. The user can enter one or more values preceding the primary unit description. Each value must be an integer, a space, or a tab. For example, the user may elect to begin each line with a counter (1, 2, 3, . . .). A counter will be used in the illustration to follow.

Format of Secondary Unit Information

Regardless of the software used to create the secondary unit file (a text file, Access table, or Excel spreadsheet), each line in the this file must contain at least three variables (fields). One possible format is the following:

MAINE 100 10

Explanation:

MAINE - This is a description of a secondary unit that was sampled. The description may be up to 25 characters in length and may include spaces, commas, and other special characters.

100 - This is the universe size of third-stage items within this secondary unit.

10 - This is the sample size of third-stage items within this secondary unit.

Secondary Unit File Restrictions

1. The secondary unit description, the universe size of third-stage items, and the number of sampled third-stage items must be the final three entries in each line.

2. The user can enter one or more values preceding the secondary unit description. Each value must be an integer, a space, or a tab. For example, the user may elect to begin each line with a counter (1, 2, 3, . . .), a value identifying the primary unit number in the second column, and a value identifying the secondary unit number within each primary unit in the third column. A counter and primary/secondary unit identifiers will be used in the illustration to follow.

Format of Sample Data Information

Regardless of the software used to create the primary unit file (a text file, Access table, or Excel spreadsheet), each line in the this file must contain at least one variable (field) if each line in the sample data file contains a single piece of information or at least two variables (fields) if each line in the sample data file contains two pieces of information. One possible format is the following:

120.34 85.50

Explanation:

120.34 - This is the quantity being reviewed by the user. The number, for example, could be the value of a voucher or the quantity of items on an inventory card. If the quantity is negative, then a minus sign must precede the number. The user must insert at least one space between numeric values.

85.50 - 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 $120.34 claimed by a vendor only $34.84 was actually owed, then the difference amount entered would be $85.50. The user must insert at least one space between the numeric values.

The user must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Sample Data File Restrictions

1. If each line contains a single piece of information (e.g., a difference value), this value must be the final entry. If each line contains two pieces of information (e.g., an examined amount and an audited amount), theses values must be the final two entries.

2. The user can enter one or more values preceding the sample data value(s). Each value must be an integer, a space, or a tab. For example, the user may elect to begin each line with a counter (1, 2, 3, . . .), a value identifying the primary unit number in the second column, a value identifying the secondary unit number within each primary unit in the third column, and a value identifying the third-stage unit number within each sampled primary/secondary unit in the fourth column. A counter and primary/secondary/third-stage unit identifiers will be used in the illustration to follow.

Input From Text Files

The three data files 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).

Sample Data File

The sample data file in this illustration (C:\TEMP\DATA3STG.TXT) consists of the examined amounts for three primary units (REGION 1, REGION 2, AND REGION 3) with 2, 1, and 2 secondary units, respectively, selected from the three sampled primary units. Ten third-stage units are sampled from each selected secondary unit.

The sample data file used here (C:\TEMP\DATA3STG.TXT) is shown next. Notice that this file contains a counter (1, 2, 3, . . .) in the first column, a value identifying the primary unit number in the second column, a value identifying the secondary unit number within each primary unit in the third column, and a value identifying the third-stage unit number within each sampled

primary/secondary unit in the fourth column. These columns are not required but are recommended for ease of navigating through this file.

```
1     1    1    1     103
2     1    1    2     97
3     1    1    3     94
4     1    1    4     100
5     1    1    5     93
6     1    1    6     102
7     1    1    7     104
8     1    1    8     104
9     1    1    9     103
10    1    1    10    100
11    1    2    1     152
12    1    2    2     152
13    1    2    3     147
14    1    2    4     161
15    1    2    5     144
16    1    2    6     153
17    1    2    7     154
18    1    2    8     151
19    1    2    9     148
20    1    2    10    158
21    2    1    1     167
22    2    1    2     178
23    2    1    3     182
24    2    1    4     177
25    2    1    5     183
26    2    1    6     181
27    2    1    7     180
28    2    1    8     180
29    2    1    9     181
30    2    1    10    191
31    3    1    1     214
32    3    1    2     206
33    3    1    3     228
34    3    1    4     201
35    3    1    5     206
36    3    1    6     211
37    3    1    7     194
38    3    1    8     206
39    3    1    9     224
40    3    1    10    210
41    3    2    1     240
42    3    2    2     249
43    3    2    3     246
44    3    2    4     254
```

```
45    3    2    5    247
46    3    2    6    253
47    3    2    7    253
48    3    2    8    245
49    3    2    9    248
50    3    2    10   255
```

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

Primary Unit File

The primary unit file used here (C:\TEMP\PRIMARY3STG.TXT) is shown below. Notice that this file contains a counter (1, 2, 3, . . .) in the first column. This column is not required.

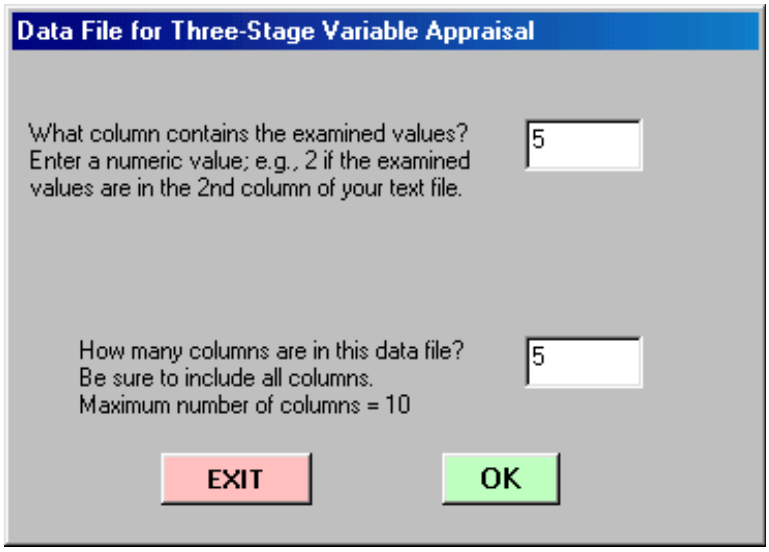
```
1  Region 1  10  2
2  Region 2   8  1
3  Region 3  15  2
```

Secondary Unit File

The secondary unit file used here (C:\TEMP\SECONDARY3STG.TXT) is shown next. Notice that this file contains a counter (1, 2, 3, . . .) in the first column, a value identifying the primary unit number in the second column, and a value identifying the secondary unit number within each primary unit in the third column. These columns are not required but are recommended for ease of navigating through this file.

```
1  1  1  Maine       100   10
2  1  2  Vermont     125   10
3  2  1  New York     80   10
4  3  1  Texas       140   10
5  3  2  Arkansas     85   10
```

After opening the three text files, the program will return to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input data and the total number of columns in the sample data file using the screen below.



After entering the column information, click on **OK**. The program will resume processing.

Input From an Access Database

The necessary information may be stored in three tables within the same 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.

Primary Unit Table

The primary unit table used here (Primary3Stg) is shown below. Notice that this file contains a counter (1, 2, 3, . . .) in the first field. This field is not required.

Primary3Stg : Table

Primary Unit Sample ID	Primary Unit Description	Secondary Universe	Secondary Sample
1	Region 1	10	2
2	Region 2	8	1
3	Region 3	15	2

Secondary Unit Table

The secondary unit table used here (Secondary3Stg) is shown below. Notice that this file contains a counter (1, 2, 3, . . .) in the first field, a value identifying the primary unit number in the second field, and a value identifying the secondary unit number within each primary unit in the third field. These fields are not required but are recommended for ease of navigating through this file.

Secondary3Stg : Table

Unique Identifier	PSU Sample ID	SSU Sample ID	SSU Description	3rd Stage Universe	3rd Stage Sample
1	1	1	Maine	100	10
2	1	2	Vermont	125	10
3	2	1	New York	80	10
4	3	1	Texas	140	10
5	3	2	Arkansas	85	10

Sample Unit Table

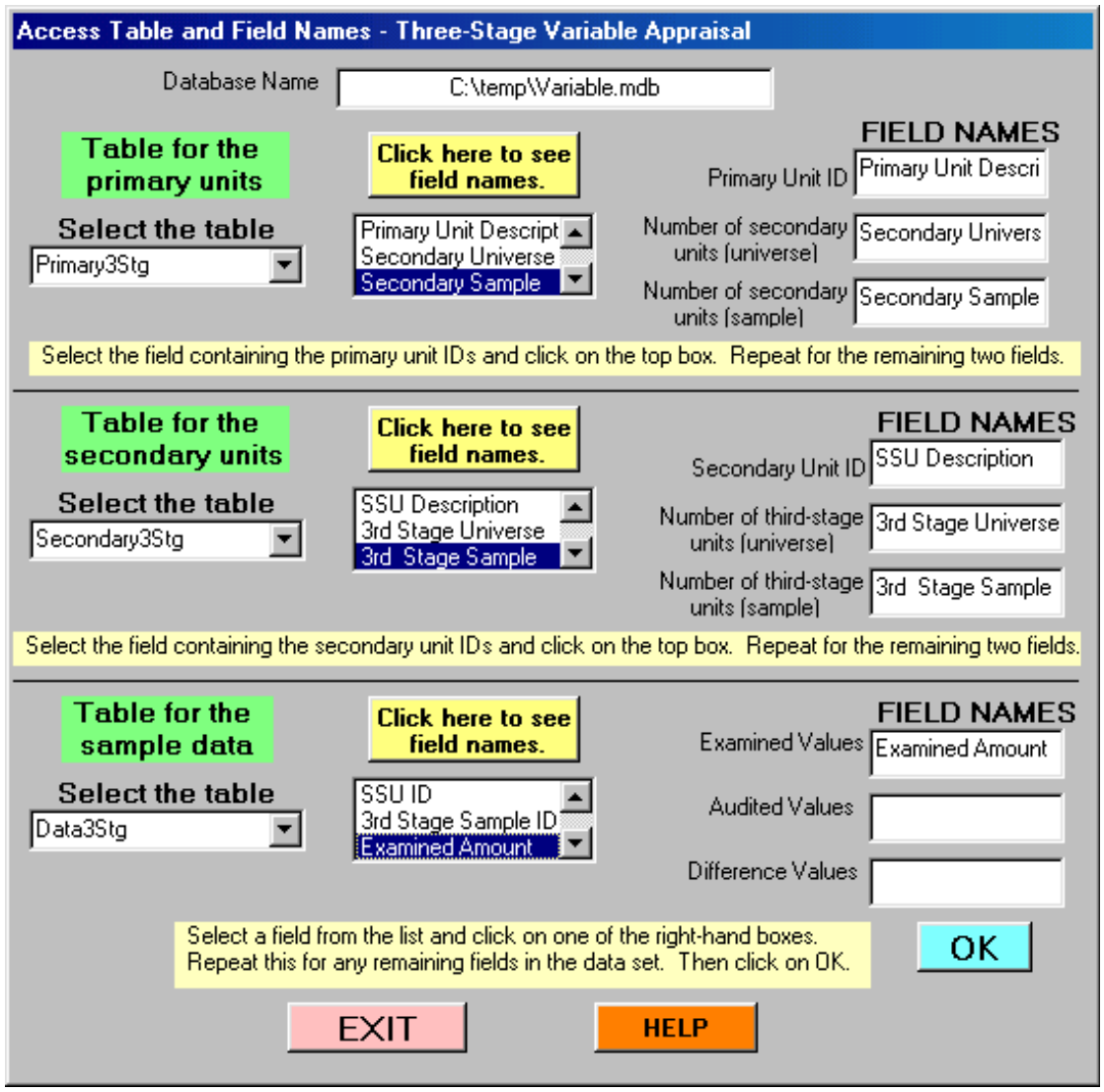
The first 10 rows of the data table used here (Data3Stg) is shown below. Notice that this file contains a counter (1, 2, 3, . . .) in the first field, a value identifying the primary unit number in the second field, a value identifying the secondary unit number within each primary unit in the third field, and a value identifying the third-stage unit number within each sampled primary/secondary unit in the fourth field. These fields are not required but are recommended for ease of navigating through this file.

Unique ID	PSU ID	SSU ID	3rd Stage Sample ID	Examined Amount
1	1	1	1	103
2	1	1	2	97
3	1	1	3	94
4	1	1	4	100
5	1	1	5	93
6	1	1	6	102
7	1	1	7	104
8	1	1	8	104
9	1	1	9	103
10	1	1	10	100

After opening the database, the user will be asked to select the name of the table containing the primary unit information, the table containing the secondary unit information, and the table containing the sample data using the following form. For the primary unit table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (Primary3Stg for this illustration), click on **Click here to see field names**. This list will contain the field names for this table. To select the field names, click on the field name for the field containing the primary unit identification ("Primary Unit Description" in this illustration) and click on the box labeled "Primary Unit ID." The field name will appear in this box. Repeat this procedure for the two fields containing the universe and sample sizes.

For the secondary unit table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (Secondary3Stg for this illustration), click on **Click here to see field names**. This list will contain the field names for this table. To select the field names, click on the field name for the field containing the secondary unit identification ("SSU Description" in this illustration) and click on the box labeled "Secondary Unit ID." The field name will appear in this box. Repeat this procedure for the two fields containing the universe and sample sizes for each secondary unit.

To select the field name(s) for the sample data file, first select the table (Data3Stg in this illustration), click on **Click here to see 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 box labeled "Examined." The field name will then 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 then return to the input screen. **NOTE**: When the user returns to the input screen, the data file format (Examined Values for this illustration) will be selected, based on responses within the following Select Access Table and Field Names form. When using an Access input file, the user cannot change this data file format option after returning to the input screen.



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

Input From Excel Spreadsheets

With this option, the primary unit information, the secondary unit information, and the sample data must be stored in three Excel spreadsheets within the same Excel file. Select the name of the file containing the spreadsheets in the preceding **Open Data File** step. This file must have

the standard Excel extension (.XLS). For this illustration, Excel file DATA3STG.XLS will be used.

Primary Unit Spreadsheet

The contents of the spreadsheet containing the primary unit information (Primary_Info in this illustration) is shown below. Notice that this file contains a counter (1, 2, 3, . . .) in the first column. This column is not required.

	A	B	C	D
1	Primary Unit Sample ID	Primary Unit Description	Secondary Universe	Secondary Sample Size
2	1	Region 1	10	2
3	2	Region 2	8	1
4	3	Region 3	15	2

Secondary Units Spreadsheet

The contents of the spreadsheet containing the secondary unit information (Secondary_Info in this illustration) are shown next. Notice that this file contains a counter (1, 2, 3, . . .) in the first column, a value identifying the primary unit number in the second column, and a value identifying the secondary unit number within each primary unit in the third column. These columns are not required but are recommended for ease of navigating through this file.

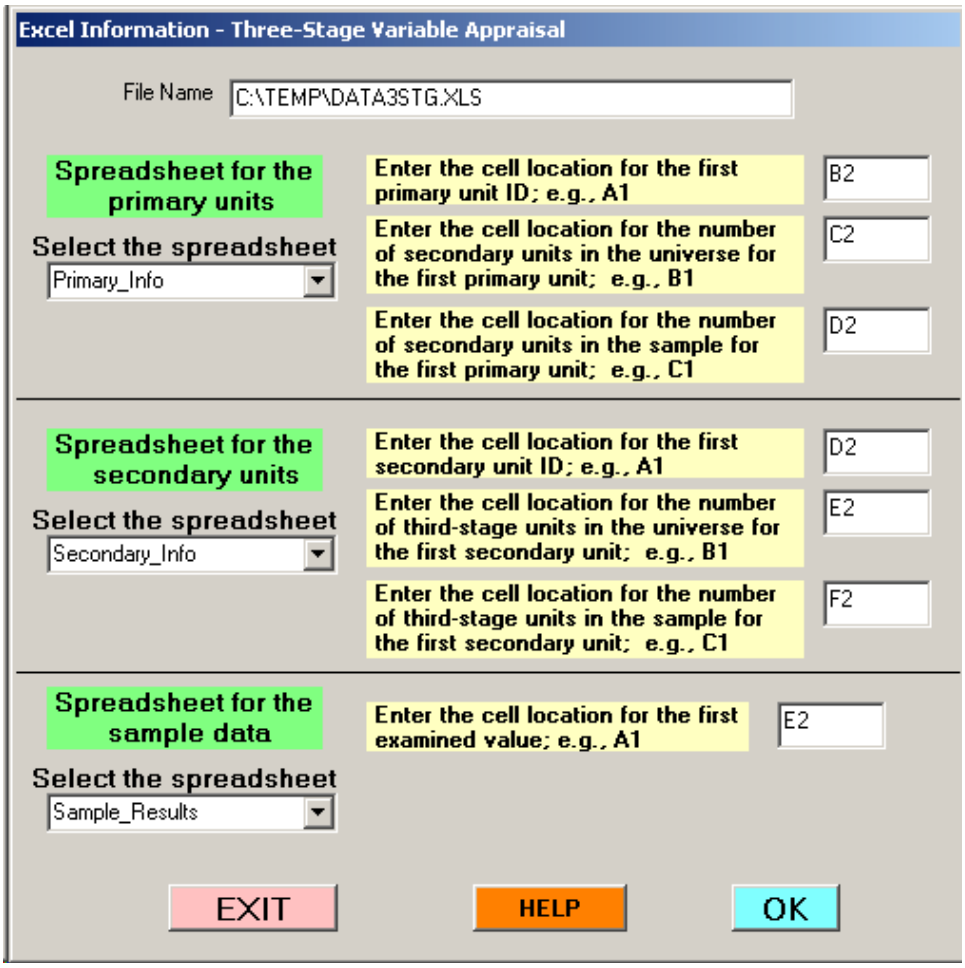
	A	B	C	D	E	F
1	Unique Identifier	Primary Unit Sample ID	Secondary Unit Sample ID	Secondary Unit Description	3rd Stage Universe	3rd Stage Sample
2	1	1	1	Maine	100	10
3	2	1	2	Vermont	125	10
4	3	2	1	New York	80	10
5	4	3	1	Texas	140	10
6	5	3	2	Arkansas	85	10

Sample Data Spreadsheet

The first 10 observations in the sample data table (Sample_Results) are shown next. Notice that this file contains a counter (1, 2, 3, . . .) in the first column, a value identifying the primary unit number in the second column, a value identifying the secondary unit number within each primary unit in the third column, and a value identifying the third-stage unit number within each sampled primary/secondary unit in the fourth column. These columns are not required but are recommended for ease of navigating through this file.

	A	B	C	D	E
1	Unique ID	Primary Unit Sample ID	Secondary Unit Sample	Third Stage Sample ID	Examined Amount
2	1	1	1	1	103
3	2	1	1	2	97
4	3	1	1	3	94
5	4	1	1	4	100
6	5	1	1	5	93
7	6	1	1	6	102
8	7	1	1	7	104
9	8	1	1	8	104
10	9	1	1	9	103
11	10	1	1	10	100

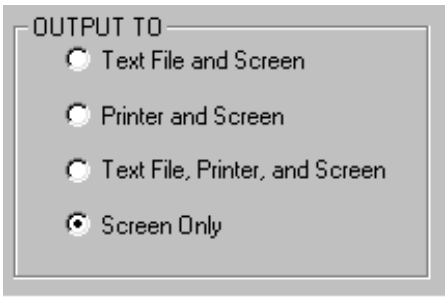
After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will be displayed. For this illustration, the various boxes should be filled in as shown. After entering the cell locations, click on **OK**. The program will resume processing.



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 "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.

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**.

Program Output

If the complete option was selected by the user, the following information will be displayed for the examined, adjusted, and difference sections of the output. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

FIRST STAGE / SECOND STAGE	The description of the first- or second-stage unit as shown in the sample data file.
SAMPLE SIZE	The number of items sampled in the particular second stage.
NONZEROES	The quantity of sample items in the particular second stage that had values other than zero.
SAMPLE 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.
VARIANCE	A measurement of the variation of the sampled items about the sample mean.
UNIVERSE	The quantity of the third-stage items within each sampled second-stage unit from which the sample was drawn.
POINT ESTIMATE	The single estimate for the universe total of the second stage unit based on the sample mean and universe.

If the user selects the summary option the results will be displayed only at the first-stage level. The universe of third-stage items and the variance at each second stage are not included in the summary output.

The following information will be displayed in both the complete and summary output:

UNIVERSE	For each stage, the number of items in the respective universes. For the second and third-stages, the universes represent the total for the prior-stage sample. For example, if 10 primary units are sampled from a universe of 100 units, then the secondary universe is the total of secondary items in the 10 primary units reviewed and not the total of secondary units in the 100 primary units.
SAMPLED	The number of units sampled at each stage.
OVERALL POINT ESTIMATE	The estimate of the universe total.

OVERALL STANDARD ERROR	A measurement of the variation of the overall point estimate of the universe total with respect to all possible point estimates for this universe and these sample sizes.
CONFIDENCE LEVELS	The confidence associated with the ability of 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 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	The result of dividing the precision amount by the point estimate.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUT3STG.TXT, shown immediately following. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004  THREE-STAGE UNRESTRICTED VARIABLE APPRAISAL      Time: 17:03
                        AUDIT/REVIEW: Variable 3-Stage

                     DATA FILE USED: C:\temp\Data3stg.TXT

       ----------------------- E X A M I N E D -----------------------
FIRST STAGE      SAMPLE  NON-
  SECOND STAGE    SIZE  ZEROES SAMPLE MEAN VARIANCE   UNIVERSE   POINT ESTIMATE
================ ====== ====== =========== ========  ==========  ==============
REGION 1
  MAINE             10     10      100.00    16.44         100          10,000
  VERMONT           10     10      152.00    25.33         125          19,000

    COMBINED        20             14,500                   10         145,000

REGION 2
  NEW YORK          10     10      180.00    35.33          80          14,400

    COMBINED        10             14,400                    8         115,200

REGION 3
  TEXAS             10     10      210.00   102.44         140          29,400
  ARKANSAS          10     10      249.00    22.67          85          21,165

    COMBINED        20             25,283                   15         379,238


                 STAGES                UNIVERSE         SAMPLED
                   FIRST                     15               3
                   SECOND                    33               5
                   THIRD                    530              50

OVERALL POINT ESTIMATE                                                3,197,188
OVERALL STANDARD ERROR                                                1,131,084

                                        CONFIDENCE LIMITS
                                       80% CONFIDENCE LEVEL
           LOWER LIMIT                           1,747,646
           UPPER LIMIT                           4,646,729
           PRECISION AMOUNT                      1,449,542
           PRECISION PERCENT                         45.34%
           Z-VALUE USED                     1.281551565545

                                       90% CONFIDENCE LEVEL
           LOWER LIMIT                           1,336,721
           UPPER LIMIT                           5,057,654
           PRECISION AMOUNT                      1,860,467
           PRECISION PERCENT                         58.19%
           Z-VALUE USED                     1.644853626951
```

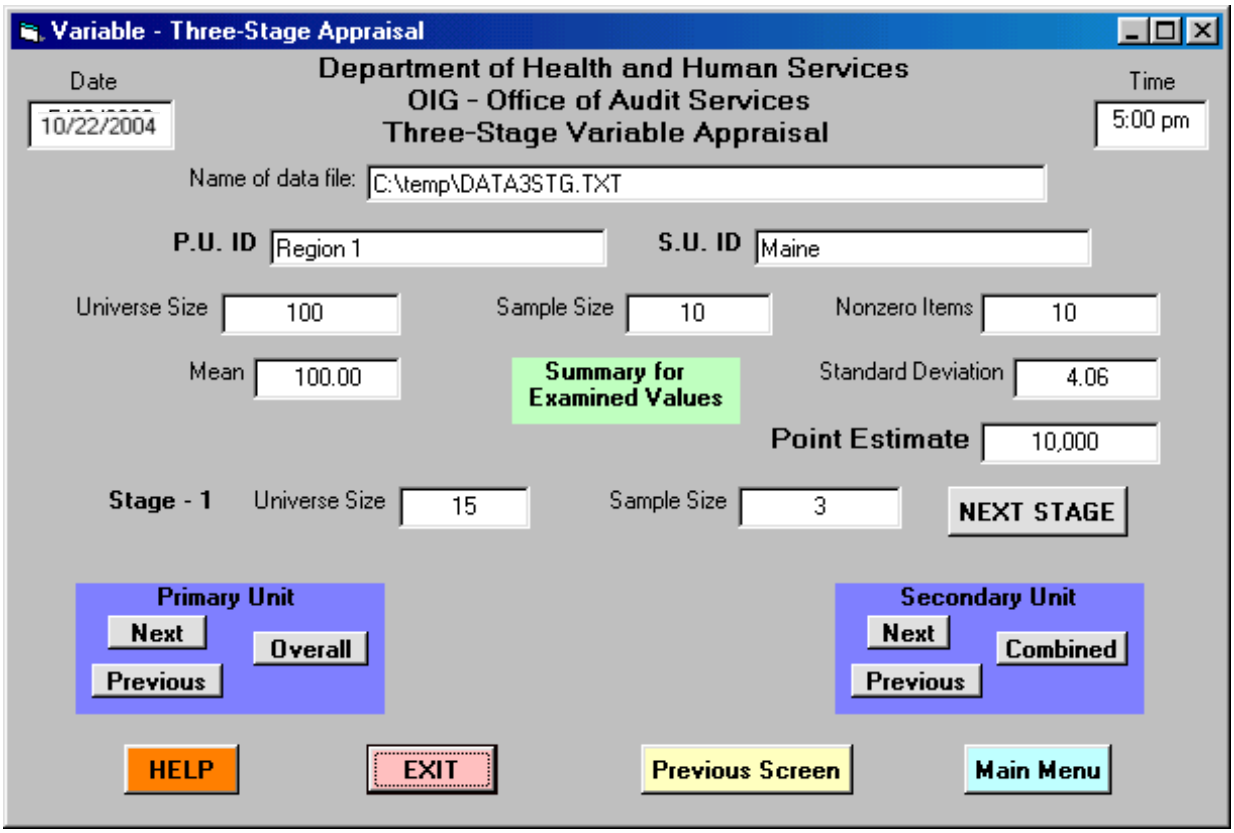
```
                                  95% CONFIDENCE LEVEL
LOWER LIMIT                                 980,304
UPPER LIMIT                               5,414,071
PRECISION AMOUNT                          2,216,883
PRECISION PERCENT                             69.34%
Z-VALUE USED                        1.959963984540
```

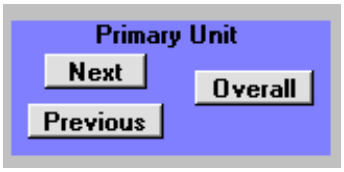
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

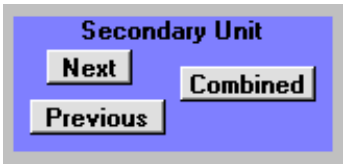
Output to Screen

The program always concludes with a summary on the screen, even if the output option selected is a text file or printer. The screen below is the summary of the examined values for the first primary unit and the first secondary unit (REGION 1, MAINE) in this illustration. If the user created a sample 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 form. For this illustration, the **Additional Summary Info** button is not visible since only the difference values were used in the sample data file.

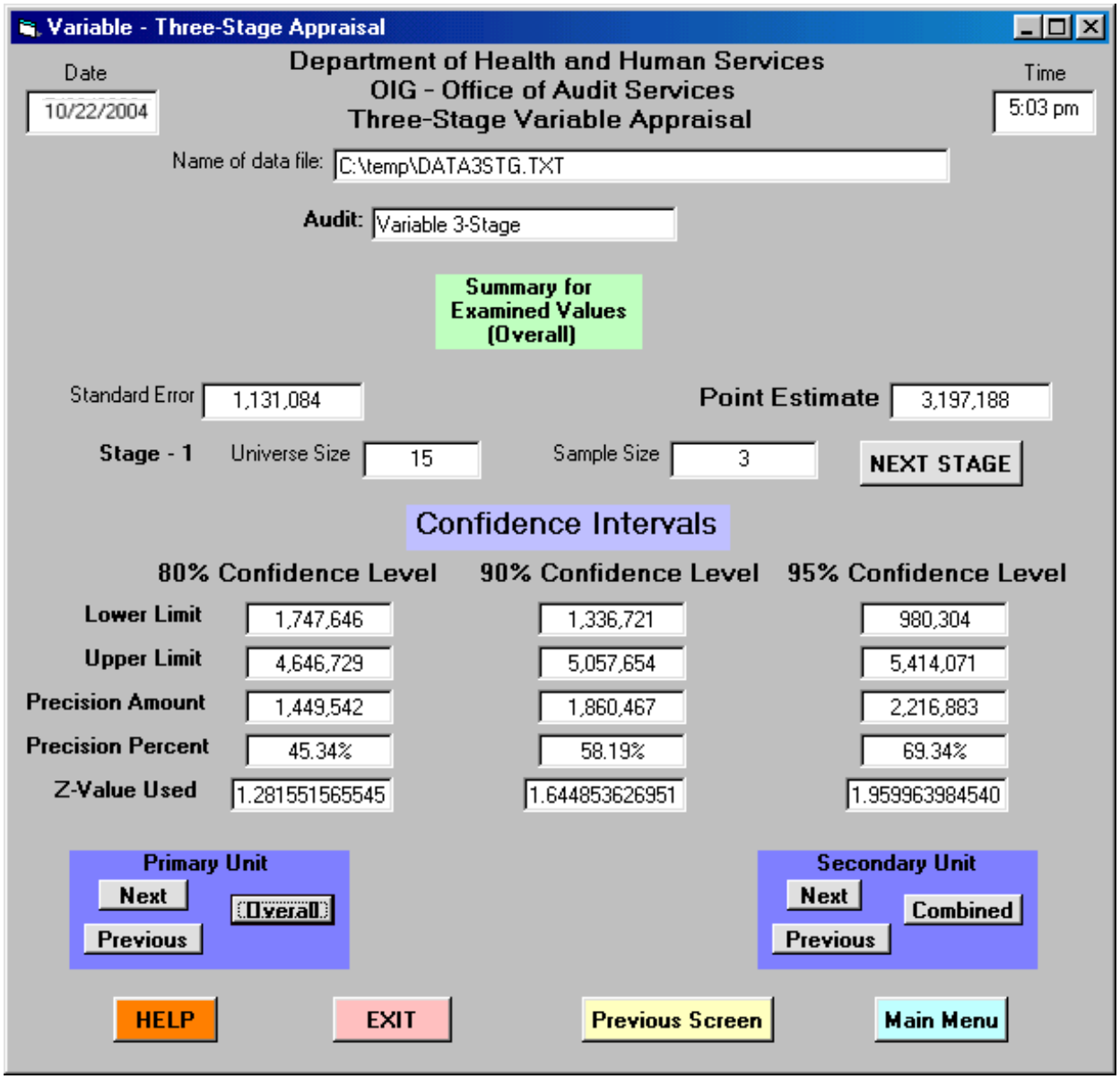
By clicking on **NEXT STAGE**, the user can see the number of primary units (Stage-1) in the universe and in the sample, secondary units (Stage-2) in the universe and in the sample, and third-stage units (Stage-3) in the universe and in the sample. Using the screen below, for this illustration, there are 15 primary units in the universe and 3 in the sample. Based on the preceding output to text file and printer, for Stage-2 the values would be 33 (universe) and 5 (sample). For Stage-3, the values would be 530 (universe) and 50 (sample).







To view summary information for the next sampled secondary unit within the selected primary unit, click on **Next** in the **Secondary Unit** box. By clicking on **Next** and **Previous**, the user can navigate through the secondary units for the selected primary unit. By clicking on **Combined,** the screen will contain the summary for the selected primary unit. Similarly, by clicking on **Next** and **Previous** in the **Primary Unit** box, the user can navigate through the sampled primary units. For each selected primary unit, the user can then switch to the **Secondary Unit** box and examine the results for each secondary unit within this primary unit. Finally, by clicking on **Overall** in the **Primary Unit** box, the summary appraisal results for the entire analysis are shown. This output screen is shown next.



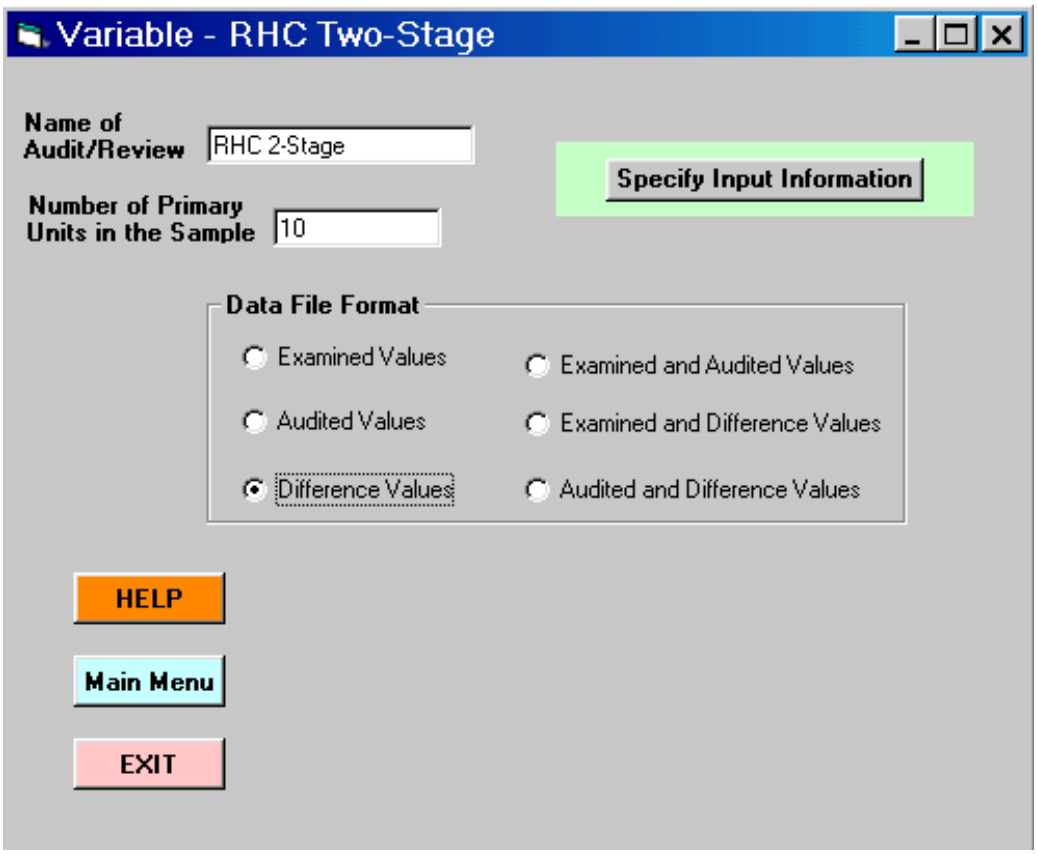
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

RHC TWO STAGE

Purpose

This program performs a two-stage variable appraisal using the Rao-Hartley-Cochran (RHC) methodology. This multistage procedure determines a confidence interval that uses approximate probability proportional to size (pps) sampling whereby the relative sizes of the sampling units are considered when selecting primary units to include in the sample. This is a variable sampling procedure allowing the user to obtain one or more quantitative 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).

Input Screen



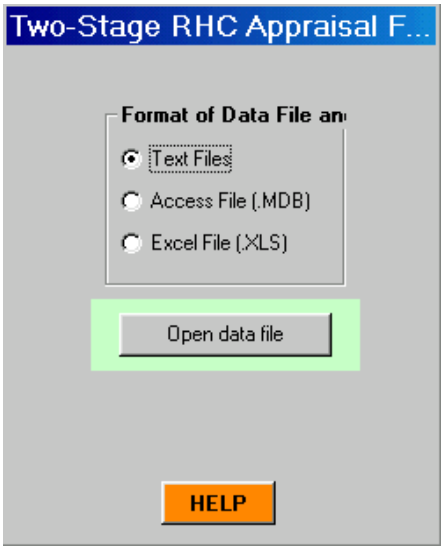
NOTE: The **Specify Input Information** button only becomes visible when the number of primary units in the sample has been entered.

Name of audit/review

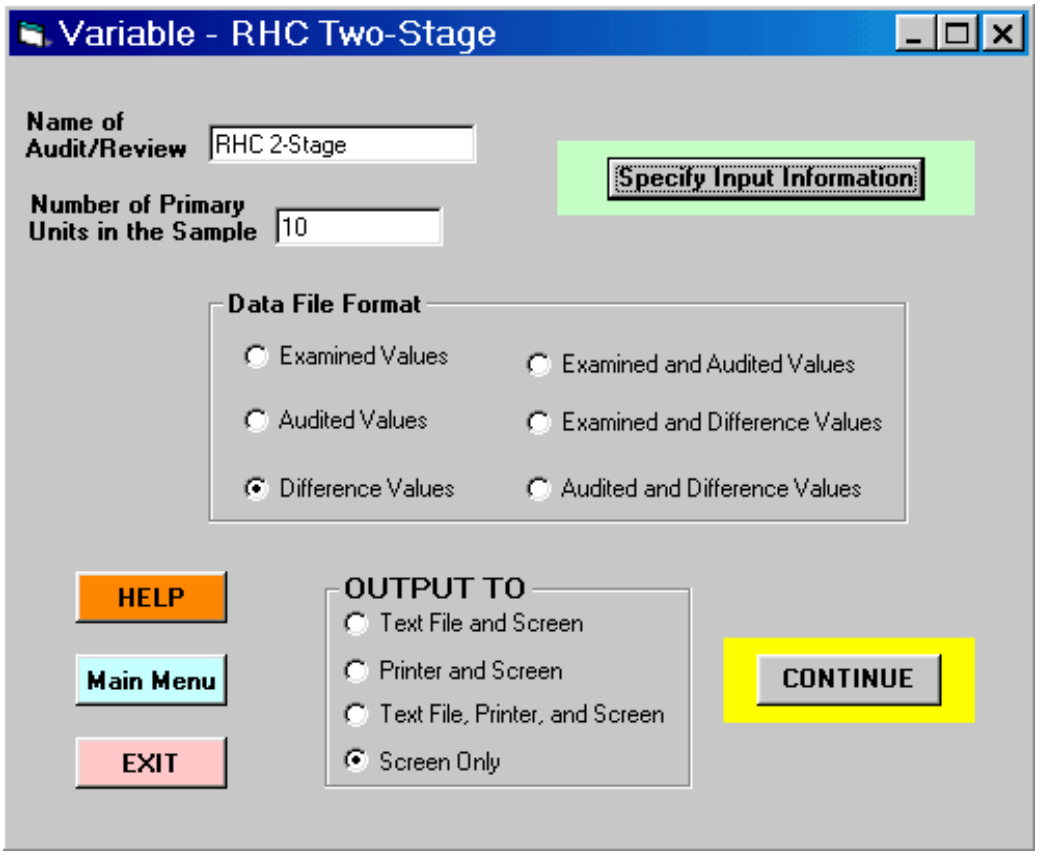
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

To enter input information, click on **Specify Input Information**. The form at the bottom of the page will appear. The sample data and primary unit (P.U.) file format can consist of two text files, two tables within the same Access database, or two spreadsheets within the same Excel file. After selecting the desired format for the sample data file, click on **Open data file**. **Note**: If the **Access File (.MDB)** option is selected, this button will change to **Open Access database** and if the **Excel file (.XLS)** option is selected, this button will change to **Open Excel file**. Using the standard Windows "Open" file screen, locate the sample data file and double-click on it.

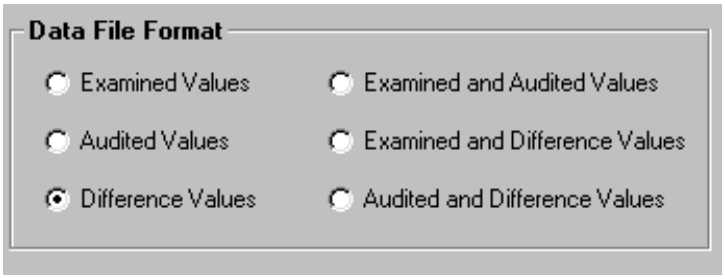
If the **Text files** option is being used, an **Open Primary Unit File** button will appear in the form below after the sample data file has been opened. The primary unit file is a slightly modified version of the file created by the RHC Sample Selection program. Click on the **Open Primary Unit File** button and, using the standard Windows "Open" file screen, locate the primary unit file and double click on it.



When all files have been opened, the full input screen will appear.



Format of Input File



Prior to executing this program, the user must create a sample 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 sample 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.

43.00 - If two pieces of information are 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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Format of Primary Unit Information

When the RHC Sample Selection program is used, a file is created containing information on the clustering of the primary units in the universe as well as the primary units selected. It may be a text file, a table within an Access database, or an Excel spreadsheet. This is an easy file to create since the user only needs to insert the sample sizes (the third column, denoted as "15" in the line below) into the file created by the RHC Sample Selection program.

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

PRIMARY #1 200 15 1,100 12,500 8

Explanation:

PRIMARY #1 - This is a description of the primary unit. A maximum of 30 characters, including spaces, will be accepted.

200 - This is the number of secondary units in the universe for this particular primary unit. The user may include commas in the value.

15 - This is the number of secondary units in the sample for this particular primary unit. The user may include commas in the value.

1,100 - This is the size factor assigned by the user to the primary unit (e.g., square footage, number of transactions, etc.). Commas and decimal points may be included in the values.

12,500 - This is the total size factor for the group in which the primary unit was randomly placed. Commas and decimal points may be included in the values.

8 - This is the quantity of primary units that were randomly placed into this group. Commas may be included in the value.

The user must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs should be placed between values so that the sample values align vertically. The program assumes one or more spaces or tabs as the only delimiters between pieces of data.

Input From Text Files

The sample data and primary unit information may be stored in a text file format. There are several ways the users 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).

Sample Data File

As an illustration, 10 universities (primary units) were selected using the RHC Sample Selection program. The file below (C:\TEMP\DATARHC2.TXT) was obtained by sampling the 10 selected universities.

1	9	⬅ **P.U. 1 starts here**
2	2	
3	9	
4	6	
5	0	
6	5	
7	7	
8	2	
9	7	
10	6	⬅ **P.U. 2 starts here**
11	0	
12	6	
13	0	
14	3	
15	4	
16	1	
17	13	
18	8	
19	0	
20	6	
21	11	
22	8	
23	8	
24	0	
25	0	⬅ **P.U. 3 starts here**
26	6	
27	0	
28	7	
29	2	
30	0	
31	9	
32	9	
33	5	
34	0	
35	5	
36	8	
37	10	
38	5	⬅ **P.U. 4 starts here**
39	8	
40	9	
41	11	
42	2	
43	8	
44	2	
45	1	
46	7	
47	3	
48	4	
49	0	
50	6	
51	1	⬅ **P.U. 5 starts here**
52	5	
53	4	
54	0	
55	6	
56	5	
57	2	
58	7	
59	2	⬅ **P.U. 6 starts here**
60	6	
61	8	
62	9	
63	7	
64	4	
65	1	
66	10	
67	11	
68	0	
69	1	
70	6	
71	1	
72	2	⬅ **P.U. 7 starts here**
73	1	
74	6	
75	1	
76	5	
77	10	
78	11	

79	9	
80	13	
81	2	
82	2	
83	7	
84	9	
85	0	
86	6	⬅ **P.U. 8 starts here**
87	5	
88	10	
89	6	
90	10	
91	2	
94	12	
92	2	
93	0	
95	5	
96	0	
97	8	
98	2	
99	7	
100	7	⬅ **P.U. 9 starts here**
101	10	
102	2	
103	6	
104	0	
105	8	
106	4	
107	0	
108	10	
109	3	
110	2	
111	5	
112	10	
113	0	
114	0	⬅ **P.U. 10 starts here**
115	0	
116	0	
117	0	
118	8	
119	9	
120	0	
121	2	
122	8	
123	4	
124	6	
125	2	
126	1	
127	8	
128	11	

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

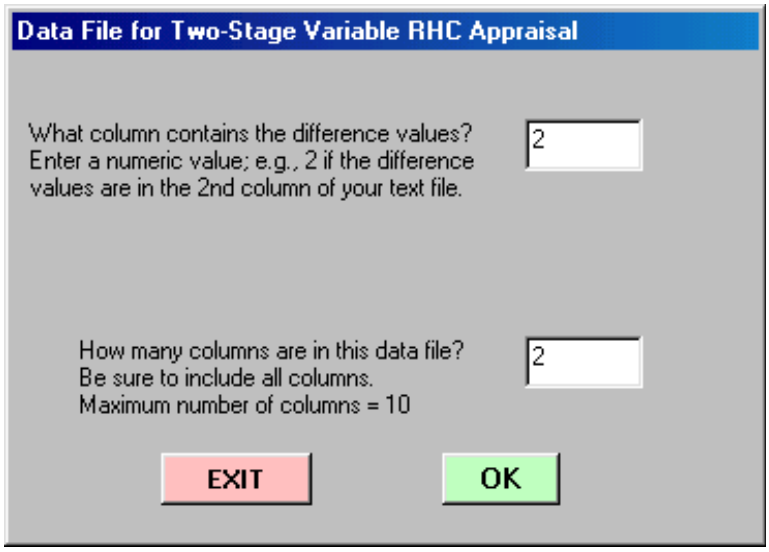
Primary Units File

For this illustration, there are 10 primary units (universities). The modified file (PRIMRHC2.TXT) created by the RHC Sample Selection program (the primary unit file) is shown below. The sample sizes (highlighted) were added to the file created by the RHC Sample Selection program.

```
UNIV21          44      9       9       86       9
UNIV62          77     15      18      105       9
UNIV61          66     13      13       81       9
UNIV82          66     13      14       94       9
UNIV52          38      8       9       81       9
UNIV11          64     13      13      104       9
```

```
UNIV66          68      14      14          96          9
UNIV55          71      14      15         107          9
UNIV90          72      14      16          95          9
UNIV86          75      15      17         101          9
```

After opening the two text files, the program will return to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input data and the total number of columns in the sample data file using the following screen:



After entering the column information, click on **OK**. The program will resume processing.

Input From an Access Database

The sample data and primary unit information must be stored in two tables within the same Access database. Select the name of the database containing the input tables 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.

Sample Data Table

For this illustration, the data will be stored in table DATARHC2 within Access database C:\TEMP\VARIABLE.MDB. The first 15 rows of Access table DATARHC2 are shown next.

DATARHC2 : Table

Counter	Difference
1	9
2	2
3	9
4	6
5	0
6	5
7	7
8	2
9	7
11	6
12	0
13	6
14	0
15	3

Primary Units Table

For this illustration, there are 10 primary units (universities). The modified table (PRIMRHC2) created by the RHC Sample Selection program (the primary unit file) is shown below. The sample sizes in the Secondary-Sample field were added to the file created by the RHC Sample Selection program.

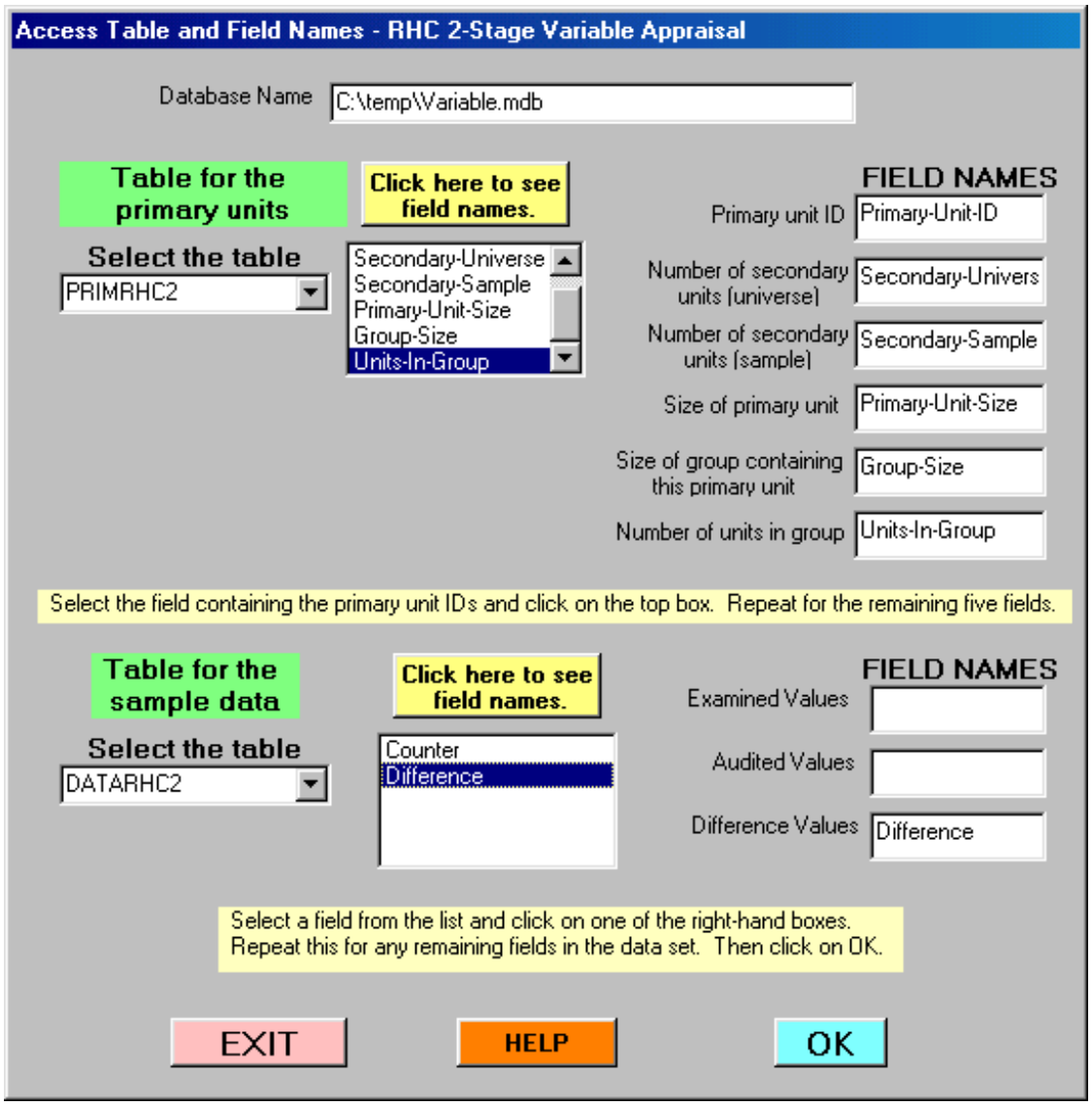
PRIMRHC2 : Table

Primary-Unit-ID	Secondary-Universe	Secondary-Sample	Primary-Unit-Size	Group-Size	Units-In-Group
UNIV21	44	9	9	86	9
UNIV62	77	15	18	105	9
UNIV61	66	13	13	81	9
UNIV82	66	13	14	94	9
UNIV52	38	8	9	81	9
UNIV11	64	13	13	104	9
UNIV66	68	14	14	96	9
UNIV55	71	14	15	107	9
UNIV90	72	14	16	95	9
UNIV86	75	15	17	101	9

Immediately after specifying the Access database, the user will be asked to select the name of the tables within the selected database using the form shown next. To specify the fields for the primary unit table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (PRIMRHC2 for this illustration), click on **Click here to see field names**. To select the field names, click on the field name for the field containing the primary unit

identification ("Primary-Unit-ID" in this illustration) and click on the box labeled "Primary unit ID." The field name will appear in this box. Repeat this procedure for the remaining five fields in this table.

To select the field name(s) for the sample data table, first select the table (DATARHC2 in this illustration) and click on the field name for the field containing the first piece of information in the input file ("Difference" in this illustration) and then click on the box labeled "Difference Values." The field name will then 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 return to the input screen. **NOTE**: When the user returns to the input screen, the data file format (Difference Values for this illustration) will be selected, based on responses within the following Access Table and Field Names form. When using an Access input file, the user cannot change this data file format option after returning to the input screen.



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

Input From Excel Spreadsheets

The sample data and primary unit information can be stored in Excel spreadsheets within the same Excel file. Select the name of the spreadsheet containing the sample data and primary unit

information in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATARHC2.XLS will be used.

Sample Data Spreadsheet

The first 15 rows of the spreadsheet containing the sample data (named Data) are shown below. This particular file contains line numbers (1, 2, 3, . . .) in column A. These line numbers are optional.

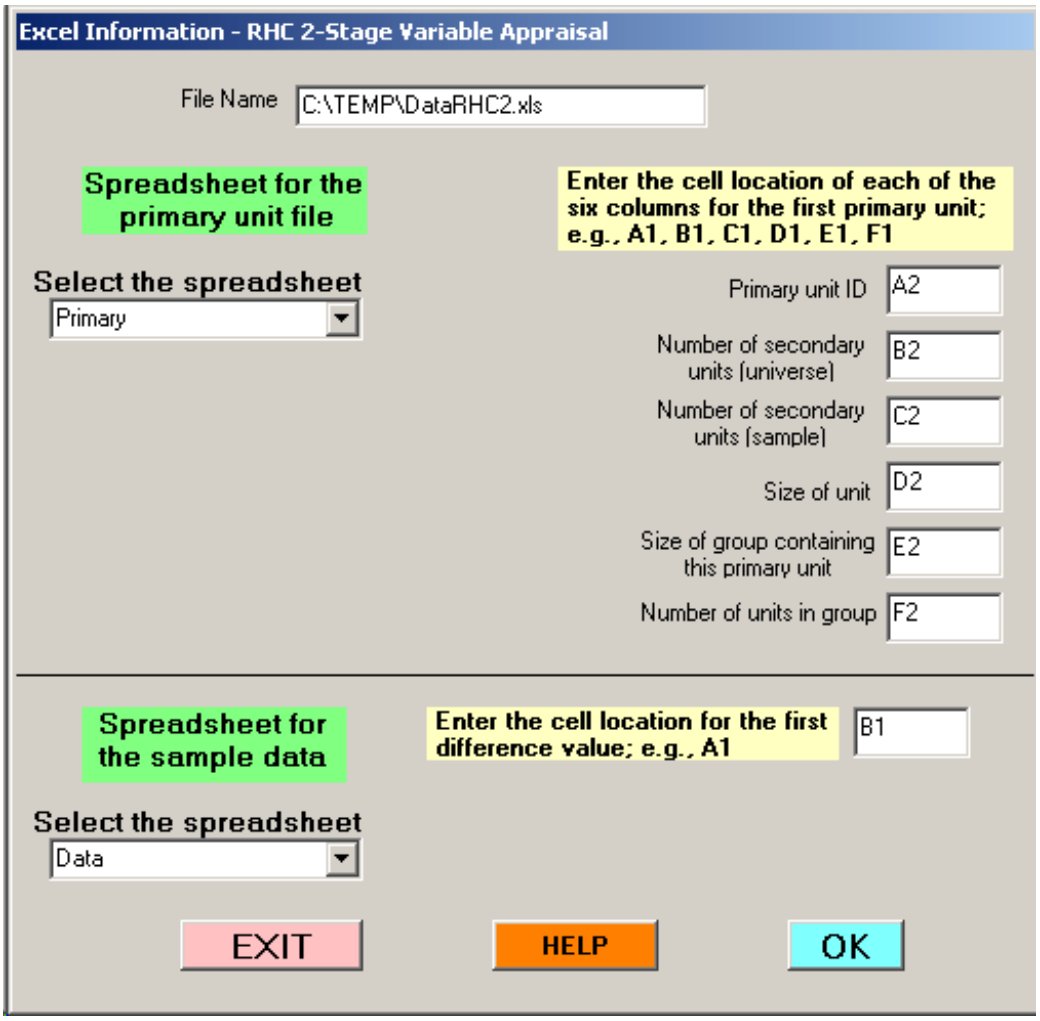
	A	B
1	1	9
2	2	2
3	3	9
4	4	6
5	5	0
6	6	5
7	7	7
8	8	2
9	9	7
10	10	6
11	11	0
12	12	6
13	13	0
14	14	3
15	15	4

Primary Units Spreadsheet

When using the Excel option, the primary unit information must be contained in another spreadsheet within the Excel file (DATARHC2.XLS in this illustration). This spreadsheet (Primary) is shown next.

	A	B	C	D	E	F
1	Primary Unit ID	Secondary Universe	Secondary Sample	Primary Unit Size	Group Size	Units in Group
2	UNIV21	44	9	9	86	9
3	UNIV62	77	15	18	105	9
4	UNIV61	66	13	13	81	9
5	UNIV82	66	13	14	94	9
6	UNIV52	38	8	9	81	9
7	UNIV11	64	13	13	104	9
8	UNIV66	68	14	14	96	9
9	UNIV55	71	14	15	107	9
10	UNIV90	72	14	16	95	9
11	UNIV86	75	15	17	101	9

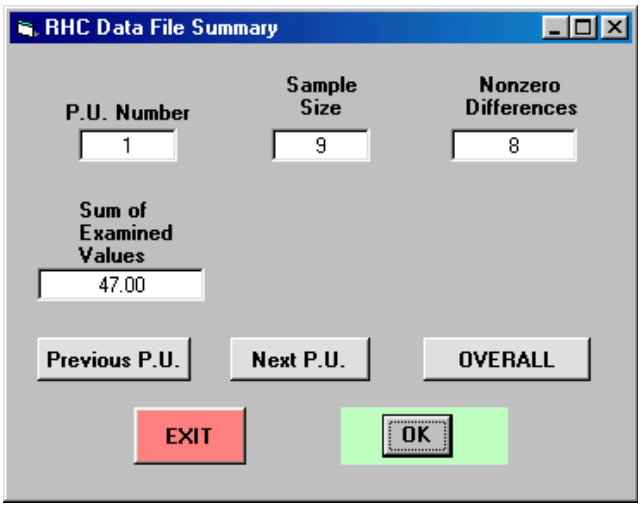
After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will be displayed. For this illustration, the various boxes should be filled in as shown. After entering the cell locations, click on **OK**. The program will resume processing.



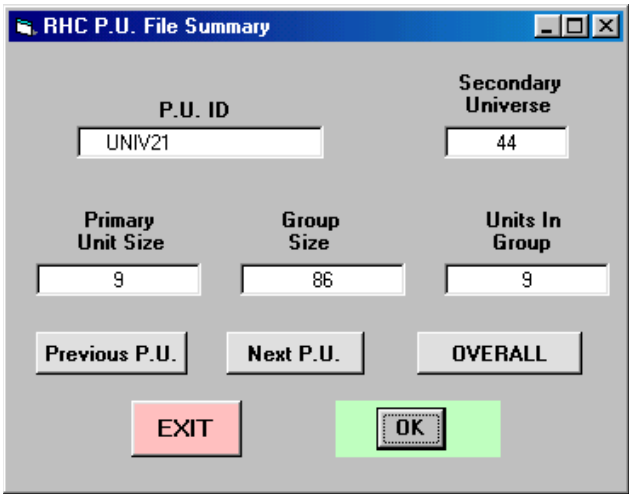
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

Summary Screens

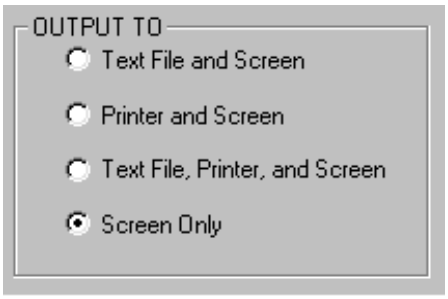
When the program has finished reading the input data files and the user clicks on **CONTINUE** in the initial screen, the summary screen will appear. The user can view the summary results for each primary unit by clicking on **Next P.U.** and **Previous P.U.**. To see the summary for all primary units combined, click on **OVERALL**.



When the user clicks on **OK** in the above screen, the screen shown below containing the first line of the primary unit file will appear. Each line of this file can be reviewed by clicking on **Next P.U.** or **Previous P.U.**. To see a summary for the primary units combined, click on **OVERALL**. To continue with the appraisal, click on **OK**.



Output Options



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or 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.

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**.

Program Output

If the complete option was selected by the user, the following information will be displayed for the examined, adjusted, and difference sections of the output. The output from this program consists of the following. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

P.U. NBR	For conserving space, the sample primary unit from each group is assigned the group number. For example, the primary unit sampled from group #1 is assigned the number 1 for identification purposes.
PRIMARY UNIT ID	The description assigned by the user to each primary unit.
SECONDARY UNIVERSE	The number of secondary units in the population of a particular primary unit.

PRIMARY UNIT SIZE	The size value selected by the user for assigning a weight to each primary unit. The weight factor could be, for example, number of employees, number of transactions, or square footage.
GROUP SIZE	The summation of the size values for all the primary units in the particular group.
UNITS IN GROUP	The number of primary units that were placed randomly in a particular group.

For each primary unit that is sampled, the following information is displayed:

SAMPLE SIZE	The number of sampled secondary units contained in this particular sampled primary unit.
SAMPLE MEAN	The average value for the appraised second-stage sample items. It is obtained by summing the items in the sample and dividing the result by the number of sample items.
SECONDARY UNIVERSE	The number of secondary units in the universe for this particular sampled primary unit. This value was originally supplied by the user in the file containing the Primary Unit information.
SIZES RATIO	The ratio of the size of the group containing this particular primary unit to the size of the primary unit itself.
POINT ESTIMATE	The estimate of the universe total for the ***group*** of primary units containing this particular primary unit. For example, suppose the primary units are split into 10 random groups, each containing 5 primary units. Suppose further that the primary unit under discussion lies in group #8. The point estimate refers to the estimate of the universe total of the five primary units in group #8. This would be repeated for the remaining sampled primary units. The point estimates would then be totaled to obtain an estimate of the universe total.
WITHIN VARIANCE	In the derivation of the standard error, the contribution of the variability of the secondary units.
BETWEEN VARIANCE	In the derivation of the standard error, the contribution of the variability of the primary units.

TOTAL VARIANCE	The sum of WITHIN VARIANCE and BETWEEN VARIANCE. The square root of this value is the STANDARD ERROR.

The results of the above information are then used in the final overall projection. The following information appears:

PRIMARY UNITS NOT SAMPLED	The number of primary units in the population minus the number of primary units in the sample.
PRIMARY UNITS IN POPULATION	The total number of primary units in the population.
POINT ESTIMATE OF POPULATION TOTAL	A single estimate for a universe value based on the summation of the point estimates for each group.
STANDARD ERROR	A measurement of the standard deviation of the estimate for the population total. It is this value that determines the width of the corresponding confidence intervals.
CONFIDENCE LEVEL	The user's level of confidence (80%, 90%, or 95%) that the actual population total will fall within the corresponding confidence interval.
LOWER LIMIT	The lower bound of the confidence interval. It is based on subtracting the precision amount from the point estimate.
UPPER LIMIT	The upper bound of the confidence interval. It is based on adding the precision amount to the point estimate.
PRECISION AMOUNT	A measurement of the closeness of the sample estimate and the corresponding population value. For a 90% confidence interval, the user would be 90% confident that the estimated population total (POINT ESTIMATE OF POPULATION TOTAL) would be within this amount of the actual value. The precision amount is calculated by multiplying the standard error by the appropriate "Z" value (Z-VALUE USED).

PRECISION PERCENT The result of dividing the precision amount by the point estimate.

Z-VALUE USED The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTRHC2.TXT, shown below. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004        RHC TWO-STAGE VARIABLE APPRAISAL            Time: 9:48
                         AUDIT/REVIEW: RHC - 2 Stage

                     DATA FILE USED: C:\temp\DATARHC2.TXT

             PRIMARY  SAMPLE   =DIFFERENCE==    NUMBER OF
              UNIT     SIZE     SAMPLE TOTAL  NONZERO ITEMS
             =======  ======   =============  =============
                1          9          47.00              8
                2         15          74.00             11
                3         13          61.00              9
                4         13          66.00             12
                5          8          30.00              7
                6         13          66.00             12
                7         14          78.00             13
                8         14          75.00             12
                9         14          67.00             11
               10         15          59.00             10

            TOTALS       128         623.00            105

                PRIMARY UNIT FILE USED: C:\temp\PRIMRHC2.txt

                                                                    UNITS
P.U.                              SECONDARY      PRIMARY                IN
NBR  PRIMARY UNIT ID               UNIVERSE     UNIT SIZE     GROUP SIZE  GROUP
==== ========================== ============= ============= ============= =====
  1  UNIV21                                44             9            86     9
  2  UNIV62                                77            18           105     9
  3  UNIV61                                66            13            81     9
  4  UNIV82                                66            14            94     9
  5  UNIV52                                38             9            81     9
  6  UNIV11                                64            13           104     9
  7  UNIV66                                68            14            96     9
  8  UNIV55                                71            15           107     9
  9  UNIV90                                72            16            95     9
 10  UNIV86                                75            17           101     9

    TOTALS                                641           138           950    90
```

DATA FILE USED: C:\temp\DATARHC2.txt

--- POINT ESTIMATES ---

P.U. NBR	SAMPLE SIZE	SAMPLE MEAN	===DIFFERENCE=== SECONDARY UNIVERSE	SIZES RATIO	POINT ESTIMATE
1	9	5.22	44	9.556	2,196
2	15	4.93	77	5.833	2,216
3	13	4.69	66	6.231	1,930
4	13	5.08	66	6.714	2,250
5	8	3.75	38	9.000	1,283
6	13	5.08	64	8.000	2,599
7	14	5.57	68	6.857	2,598
8	14	5.36	71	7.133	2,713
9	14	4.79	72	5.938	2,046
10	15	3.93	75	5.941	1,753
TOTALS:	128		641		21,582

--- VARIANCE COMPONENTS ---

P.U. NBR	WITHIN VARIANCE	BETWEEN VARIANCE	TOTAL VARIANCE
1	17,077	63,826	80,904
2	33,277	25,686	58,963
3	24,976	9,264	34,241
4	20,916	13,034	33,950
5	7,970	360,272	368,242
6	29,279	50,534	79,813
7	33,401	169,881	203,282
8	30,013	69,905	99,917
9	25,389	12,468	37,856
10	28,382	272,812	301,194
TOTALS:	250,680	1,047,683	1,298,362

DATA FILE USED: C:\temp\DATARHC2.txt

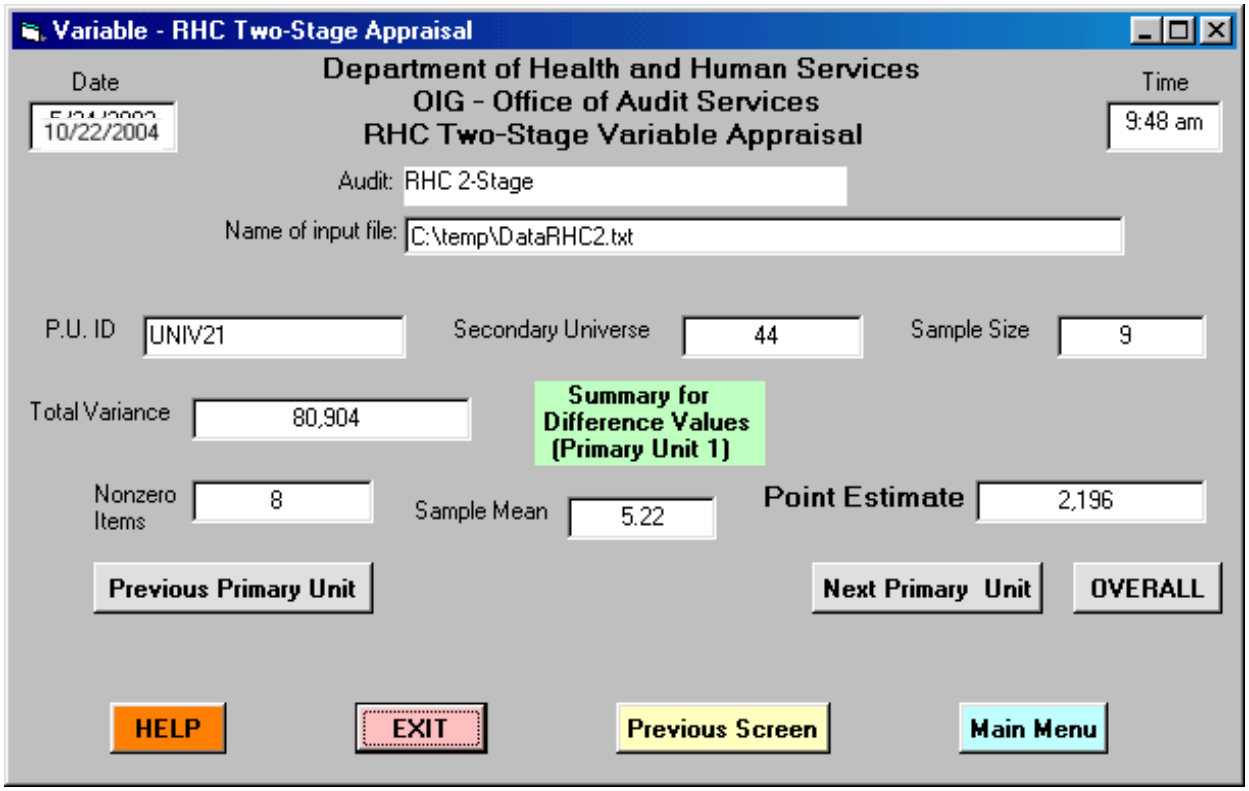
PRIMARY UNITS SAMPLED:	10
PRIMARY UNITS NOT SAMPLED:	80
PRIMARY UNITS IN POPULATION:	90
POINT ESTIMATE OF POPULATION TOTAL:	21,582
STANDARD ERROR	1,139

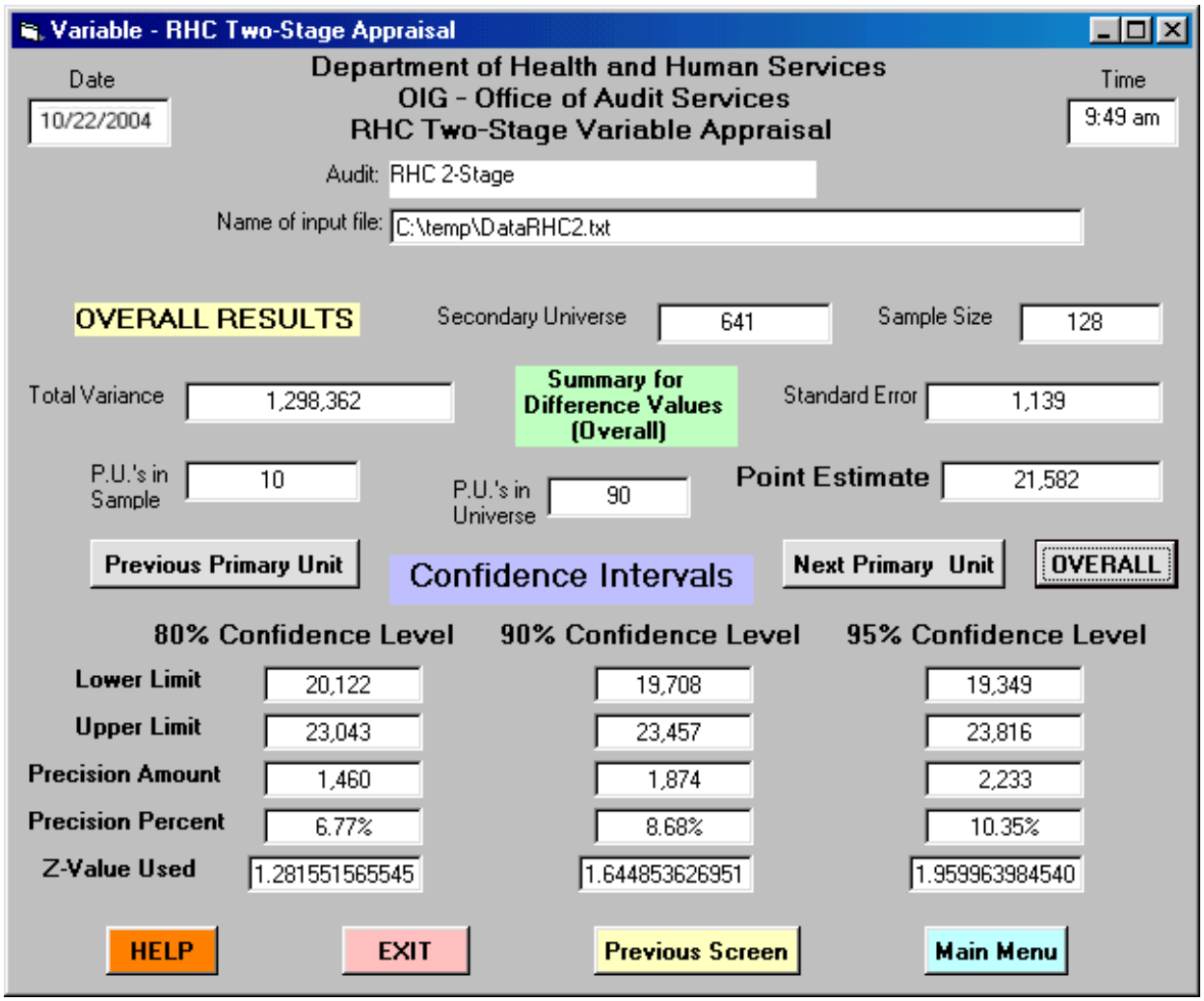
```
                                 CONFIDENCE LIMITS
                                80% CONFIDENCE LEVEL
LOWER LIMIT                                   20,122
UPPER LIMIT                                   23,043
PRECISION AMOUNT                               1,460
PRECISION PERCENT                               6.77%
Z-VALUE USED                        1.281551565545

                                90% CONFIDENCE LEVEL
LOWER LIMIT                                   19,708
UPPER LIMIT                                   23,457
PRECISION AMOUNT                               1,874
PRECISION PERCENT                               8.68%
Z-VALUE USED                        1.644853626951

                                95% CONFIDENCE LEVEL
LOWER LIMIT                                   19,349
UPPER LIMIT                                   23,816
PRECISION AMOUNT                               2,233
PRECISION PERCENT                              10.35%
Z-VALUE USED                        1.959963984540
```

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

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 screen below is the summary of the difference values for the first primary unit in this illustration. If the user created a sample 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 form. For this illustration, the **Additional Summary Info** button is not visible since only the difference values were used in the data file. To obtain the results for the second primary unit, click on **Next Primary Unit**. The user can click on **Next Primary Unit** and **Previous Primary Unit** to review the results for the individual strata. To obtain the overall results, click on **OVERALL**. The resulting summary screen immediately follows the summary for the first primary unit.





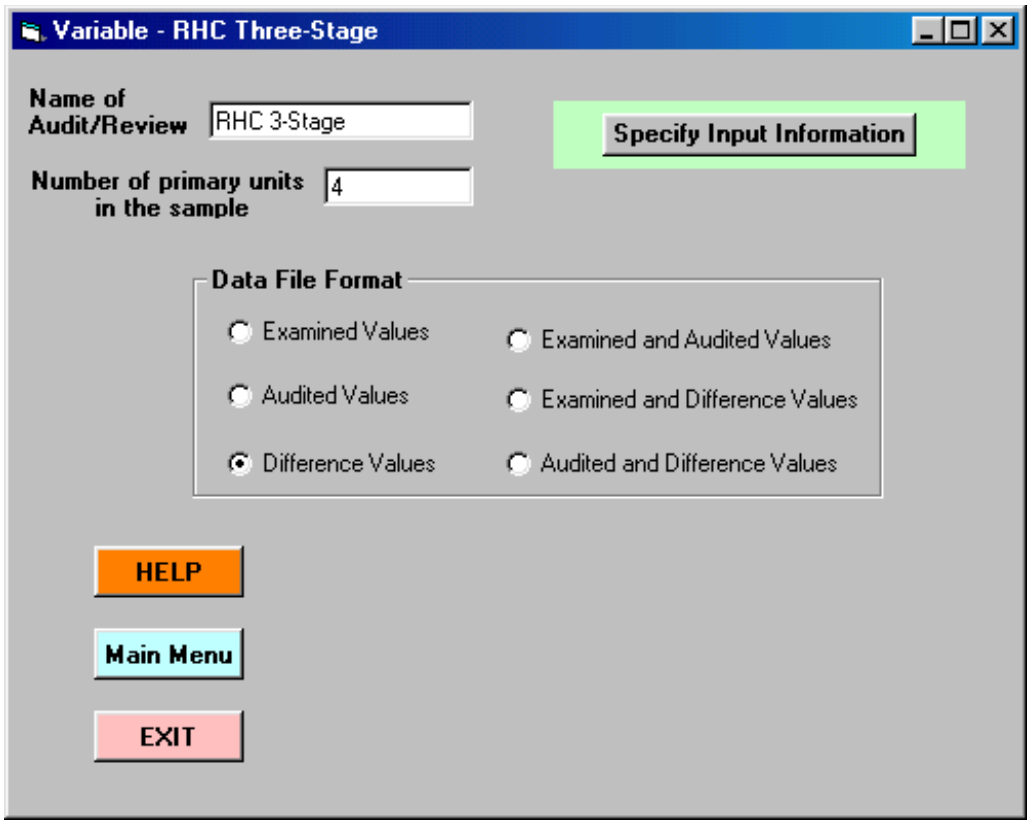
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

RHC THREE STAGE

Purpose

This program performs a three-stage variable appraisal using the Rao-Hartley-Cochran (RHC) methodology. This multistage procedure determines a confidence interval that uses approximate probability proportional to size (pps) sampling whereby the relative sizes of the sampling units are considered when selecting primary and secondary units to include in the sample. This is a variable sampling procedure allowing the user to obtain one or more quantitative 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).

Input Screen



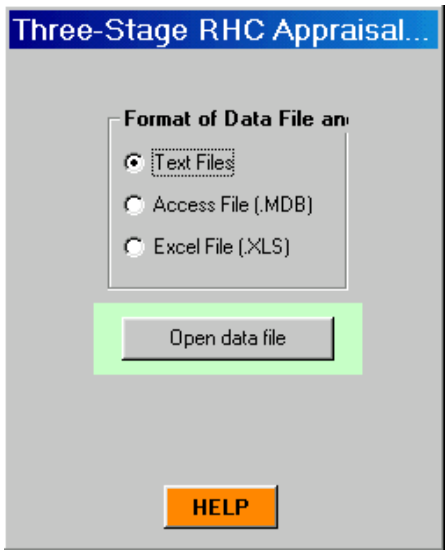
NOTE: The **Specify Input Information** button only becomes visible when the number of primary units in the sample have been entered.

Name of audit/review

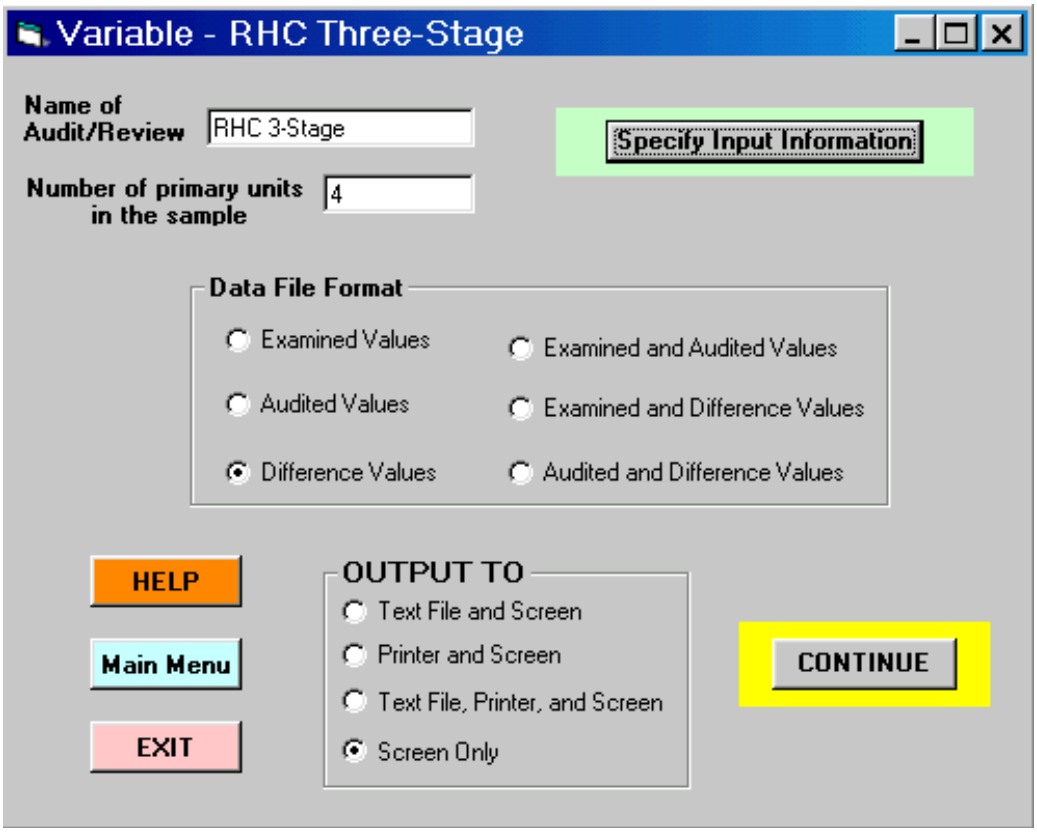
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

To enter input information, click on **Specify Input Information**. The form at the bottom of the page will appear. The sample data and primary unit/secondary unit (P.U./S.U.) file format can consist of two text files, two tables within the same Access database, or two spreadsheets within the same Excel file. After selecting the desired format for the data file, click on **Open Data File**. **Note**: If the **Access File (.MDB)** option is selected, this button will change to **Open Access database** and if the **Excel file (.XLS)** option is selected, this button will change to **Open Excel file**. Using the standard Windows "Open" file screen, locate the sample data file and double click on it.

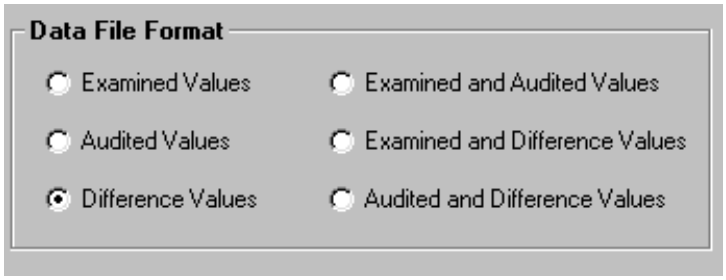
If the **Text files** option is being used, the **Open P.U./S.U. File** button will appear in the form below after the sample data file has been opened. The P.U./S.U. file is a slightly modified version of the file created by the RHC Sample Selection program. Click on the **Open P.U./S.U. File** button and, using the standard Windows "Open" file screen, locate the P.U./S.U. file and double-click on it.



When all files have been opened, the full input screen will appear.



Format of Input File



Prior to executing this program, the user must create a sample 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 sample 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.

43.00 - If two pieces of information are 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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Format of P.U./S.U. Information

When the RHC Sample Selection program is used to select the primary and secondary units, a file is created when selecting the primary units and additional files are created for the selection of the secondary units within each sampled primary unit. These files contain information on the clustering of the primary (or secondary) units in the universe as well as the primary (or secondary) units selected. These files may be text file files, tables within an Access database, or Excel spreadsheets. These files must be combined into a single file as illustrated below. It is this combined file that must be opened when clicking on the **Open P.U./S.U. File** button.

The information in this combined file must be aligned in a set format. In the P.U./S.U. file, the primary units followed by the secondary units are in the same order as in the sample data file previously discussed. The primary universe information is entered first, followed by the data for the secondary universes.

Regardless of the software used to create the P.U./S.U. file, the format should be as follows:

PRIMARY #1	**8**	**3**	**6,000**	**18,000**	**6**
SECONDARY #1	**350**	**20**	**1,000**	**1,400**	**10**

Explanation:

PRIMARY #1 - This is a brief description of the sampled primary unit. The description should be no longer than 30 characters in length.

8 - This is the quantity of secondary units in this primary unit. The user may include commas in the value.

3 - This is the number of secondary units sampled from this primary unit for this appraisal. Commas may be used in the value.

6,000 - This is the size factor assigned by the user to the primary unit (e.g., square footage, number of transactions, etc.). Commas and decimal points may be included in the values.

18,000 - This is the size factor for the group from which the primary unit was selected. Commas and decimal points may be included in the values.

6 - This is the number of primary units in the group from which this primary unit was selected. Commas may be included in the value.

SECONDARY #1 - This is a brief description of the sampled secondary unit. A maximum of 30 characters, including spaces, will be accepted.

350 - This is the quantity of third-stage units in the universe for this secondary unit. The entry may contain commas and a decimal point.

20 - This is the quantity of third-stage units in the sample for this secondary unit. The entry may contain commas and a decimal point.

1,000 - This is the secondary size factor used in weighting the secondary unit (e.g., square footage, number of transactions, etc.). Commas and a decimal point may be incorporated in the number.

1,400 - This is the size factor for the secondary group that this secondary item was sampled from. Commas and a decimal point may be used in the value.

10 - This is the number of secondary units in this group from which this secondary unit was selected. Commas may be used in the value.

The user must enter the information for primary units and secondary units on a separate line. For ease of editing, enough spaces or tabs should be placed between values so that the sample values align vertically. The program assumes one or more spaces or tabs as the only delimiters between pieces of data.

Input From Text Files

The data files may be stored in a text file format. There are several ways the users 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).

Sample Data File

As an illustration, four primary units (REGIONS 3, 7, 8, and 9) were selected from a total of 12 using the RHC Sample Selection program. Ten universities were also selected from each of the four regions using the RHC Sample Selection program. The file on the next page (C:\TEMP\DATARHC3.TXT) was obtained by sampling the 10 selected universities within each region and recording the difference values. (**NOTE**: Only a portion of the data for REGION 3 and REGION 9 is shown.)

```
1     8    ← Primary unit 1 (REGION 3), secondary unit 1 (UNIV20)
2     0
3     6
4     6
5     0
6     13
7     1
8     7    ← Primary unit 1 (REGION 3), secondary unit 2 (UNIV38)
9     2
10    13
11    13
12    4
13    6
14    0
15    15
16    12
17    9
18    0
19    13
20    10
 .
 .
 .
92    11   ← Primary unit 1 (REGION 3), secondary unit 10 (UNIV54)
93    6
94    10
95    11
96    0
97    7
98    12
```

```
99    9
100   11
 .
 .
 .
319   0    ⬅ Primary unit 4 (REGION 9), secondary unit 1 (UNIV37)
320   10
321   14
322   0
323   18
324   0
325   8
326   20
327   19
328   0
329   0
330   3
331   3
332   13   ⬅ Primary unit 4 (REGION 9), secondary unit 2 (UNIV92)
333   0
334   12
335   12
336   7
337   1
338   13
339   2
340   0
341   16
342   14
343   17
344   5
345   8
 .
 .
 .
423   19   ⬅ Primary unit 4 (REGION 9), secondary unit 10 (UNIV107)
424   17
425   13
426   12
427   13
428   12
429   11
430   14
431   13
432   0
433   1
```

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

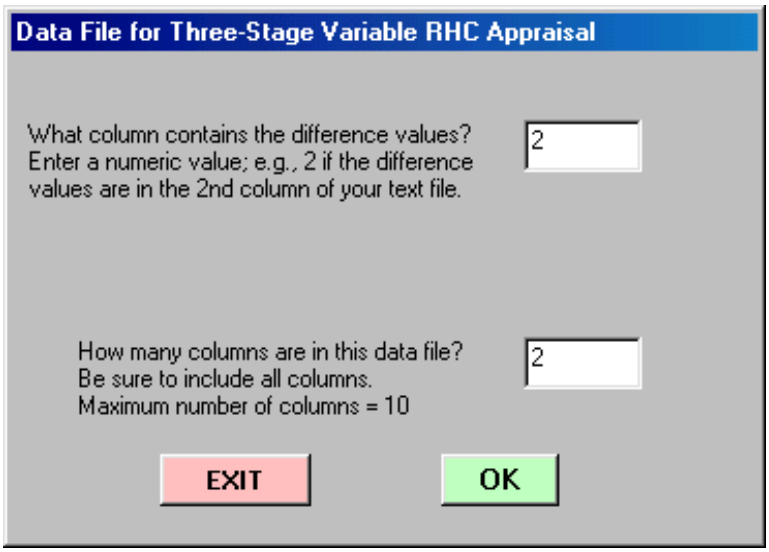
P.U./S.U. File

For this illustration, there are four primary units (regions) with 10 secondary units (universities) selected within each sampled primary unit. The file (C:\TEMP\PUSURHC3.TXT) created by combining the results of the five RHC Sample Selection programs (one for selecting the primary units and four for selecting the secondary units) is shown on the next page. The sample sizes (highlighted) were added to the files created by the five RHC Sample Selection programs.

```
REGION 3    91   10   720   3280    3
UNIV20      37    7     6     73    9
UNIV38      74   15    11     70    9
UNIV45      73   15    11     82    9
UNIV10      60   12     9     81    9
UNIV87      62   12    10     59    9
UNIV82      30    6     5     68    9
UNIV60      54   11     9     75    9
UNIV69      39    8     7     76    9
UNIV34      26    5     4     60    9
UNIV54      57   11     9     76   10
REGION 7   102   10   960   2210    3
UNIV1       56   11    10     89   10
UNIV60      56   11    10     96   10
UNIV59      67   13    13     94   10
UNIV99      80   16    14     91   10
UNIV85      67   13    13     93   10
UNIV37      31    6     6    103   10
UNIV34      42    8     8    106   10
UNIV16      53   11    10     83   10
UNIV12      66   13    13     90   11
UNIV52      60   12    11    115   11
REGION 8   118   10  1300   3710    3
UNIV19      34    7     8    137   11
UNIV104     77   15    16    122   11
UNIV66      49   10    11    122   12
UNIV110     38    8     9    137   12
UNIV83      70   14    15    121   12
UNIV14      48   10    10    147   12
UNIV78      27    5     7    122   12
UNIV105     27    5     6    144   12
UNIV12      65   13    14    122   12
UNIV112     75   15    16    126   12
REGION 9   122   10  1320   2800    3
UNIV37      64   13    14    147   12
UNIV92      73   15    15    125   12
UNIV47      71   14    15    130   12
UNIV54      70   14    15    131   12
UNIV97      56   11    12    138   12
```

```
UNIV66      76   15    16    122   12
UNIV116     50   10    10    140   12
UNIV29      33    7     8    128   12
UNIV18      26    5     7    132   13
UNIV107     55   11    11    127   13
```

After opening the two text files, the program will return to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input data and the total number of columns in the sample data file using the screen below.



After entering the column information, click on **OK**. The program will resume processing.

Input From an Access Database

The sample data and primary/secondary unit information must be stored in two tables within the same Access database. Select the name of the database containing the input tables 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 the filename and the **Open** button.

Sample Data Table

For this illustration, the sample data will be stored in table DATARHC3 within Access database C:\TEMP\VARIABLE.MDB.

The first 15 rows of Access table DATARHC3 are shown next.

DATARHC3 : Table

Counter	Difference
1	8
2	0
3	6
4	6
5	0
6	13
7	1
8	7
9	2
10	13
11	13
12	4
13	6
14	0
15	15

P.U./S.U. Table

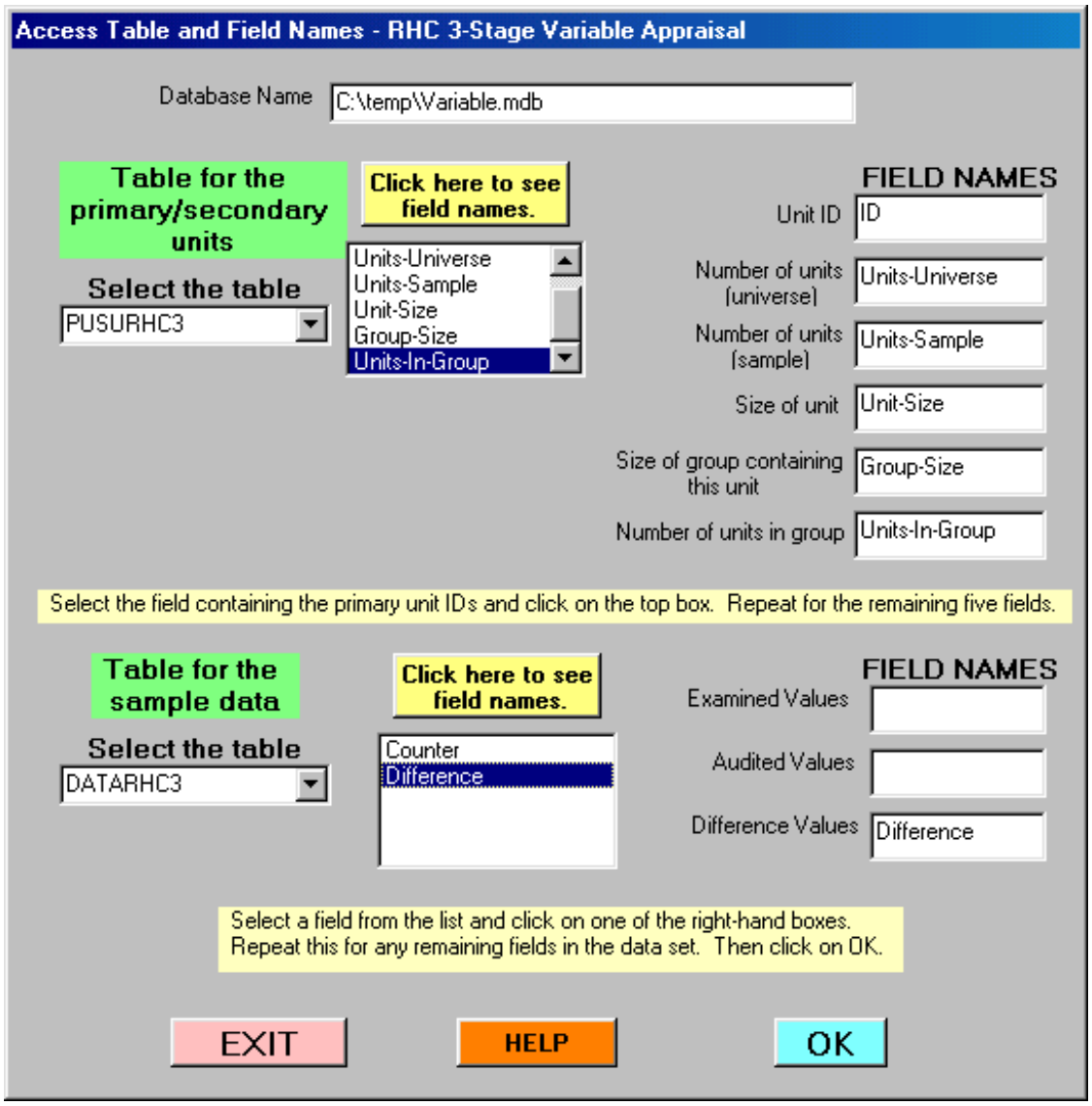
The modified table (PUSURHC3) created by combining the five passes through the RHC Sample Selection program and inserting the sample size field (Units-Sample) is shown below. This table is referred to as the **Table for the primary/secondary units**.

PUSURHC3 : Table

ID	Units-Universe	Units-Sample	Unit-Size	Group-Size	Units-In-Group
REGION3	91	10	720	3280	3
UNIV20	37	7	6	73	9
UNIV38	74	15	11	70	9
UNIV45	73	15	11	82	9
UNIV10	60	12	9	81	9
UNIV87	62	12	10	59	9
UNIV82	30	6	5	68	9
UNIV60	54	11	9	75	9
UNIV69	39	8	7	76	9
UNIV34	26	5	4	60	9
UNIV54	57	11	9	76	10
REGION7	102	10	960	2210	3
UNIV1	56	11	10	89	10
UNIV60	56	11	10	96	10
UNIV59	67	13	13	94	10

Immediately after specifying the Access database, the user will be asked to select the name of the tables within the selected database using the form on the next page. To specify the fields for the P.U./S.U. table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (PUSURHC3 for this illustration) click on **Click here to see field names**. To select the field names, click on the field name for the field containing the unit identification ("ID" in this illustration) and click on the box labeled "Unit ID." The field name will appear in this box. Repeat this procedure for the remaining five fields in this table.

To select the field name(s) for the sample data file, first select the table (DATARHC3 in this illustration) and click on the field name for the field containing the first piece of information in the input file ("Difference" in this illustration) and then click on the box labeled "Difference Values." The field name will then appear in this box. Repeat this procedure for any remaining variables of interest in this table (the only variable of interest in this illustration is the difference amount). When all the field names have been specified, click on **OK**. The program will then return to the input screen. **NOTE**: When the user returns to the input screen, the data file format (Difference Values for this illustration) will be selected, based on responses within the following Access Table and Field Names form. When using an Access input file, the user cannot change this data file format option after returning to the input screen.



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

Input From Excel Spreadsheets

The sample data and P.U./S.U. information can be stored in Excel spreadsheets within the same Excel file. Select the name of the spreadsheet containing the sample data and P.U./S.U. information in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATARHC3.XLS will be used.

Sample Data Spreadsheet

The first 15 rows of this spreadsheet (named Data) are shown below. This particular file contains line numbers (1, 2, 3, . . .) in column A. These line numbers are optional.

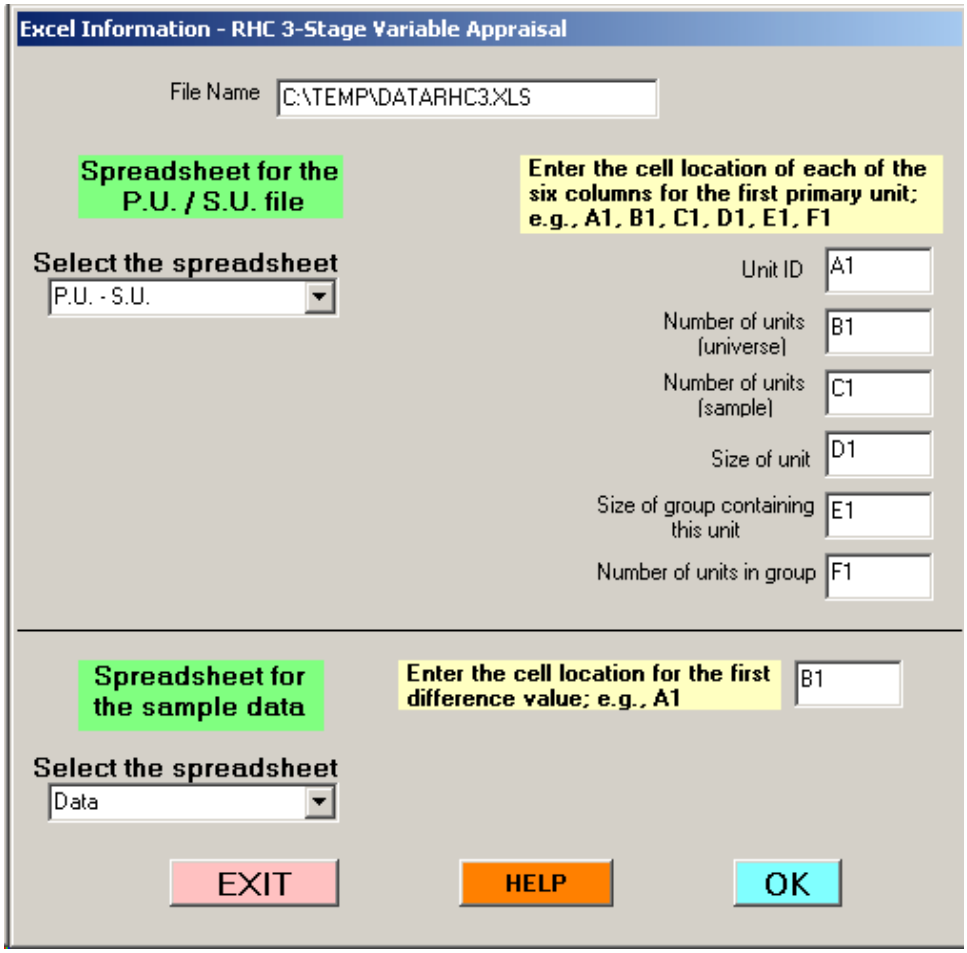
	A	B
1	1	8
2	2	0
3	3	6
4	4	6
5	5	0
6	6	13
7	7	1
8	8	7
9	9	2
10	10	13
11	11	13
12	12	4
13	13	6
14	14	0
15	15	15

P.U./S.U. Spreadsheet

When using the Excel option, the P.U./S.U. information must be contained in another spreadsheet within the Excel file (DATARHC3.XLS in this illustration). A portion of this spreadsheet (named P.U.-S.U.) is shown on the next page.

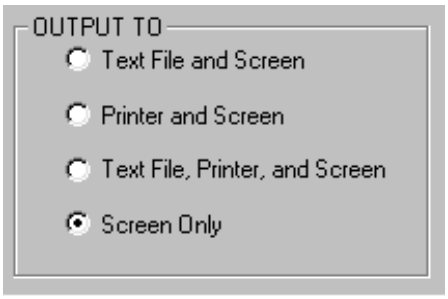
	A	B	C	D	E	F
1	REGION3	91	10	720	3280	3
2	UNIV20	37	7	6	73	9
3	UNIV38	74	15	11	70	9
4	UNIV45	73	15	11	82	9
5	UNIV10	60	12	9	81	9
6	UNIV87	62	12	10	59	9
7	UNIV82	30	6	5	68	9
8	UNIV60	54	11	9	75	9
9	UNIV69	39	8	7	76	9
10	UNIV34	26	5	4	60	9
11	UNIV54	57	11	9	76	10
12	REGION7	102	10	960	2210	3
13	UNIV1	56	11	10	89	10
14	UNIV60	56	11	10	96	10
15	UNIV59	67	13	13	94	10

After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will be displayed. For this illustration, the various boxes should be filled in as shown. After entering the cell locations, click on **OK**. The program will resume processing.



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

Output Options



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or 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.

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**.

Program Output

The following information will be displayed for the examined, adjusted, and difference sections of the output. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

The output comes in three sections. The first section provides summary information from the sample data and universe files identified by the user. The next section displays statistics for each of the secondary and primary units. In addition, the results of variance calculations are displayed. The third section displays the overall projections of the appraisal.

PRIMARY/ SECONDARY IDENTIFICATION	The description information obtained from the file of universe data identified by the user.
THIRD-STAGE UNIVERSE	The number of third-stage units that make up the population within the respective secondary unit.

SAMPLE SIZE The number of third-stage items sampled in the secondary unit.

SAMPLE VALUE The summation of the values entered for each of the third-stage sample items in the secondary unit.

NONZERO COUNT The number of third-stage sample items that had a value other than zero.

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

SIZES RATIO The ratio of the size of the group containing this particular secondary unit to the size of the secondary unit itself. The size factors are obtained from the universe file indicated by the user.

POINT ESTIMATE The estimate of the universe total for the group of secondary units containing this particular secondary unit. For example, suppose the secondary units are split into 10 random groups, each containing 5 secondary units. Suppose further that the sampled secondary unit lies in group #8. The point estimate refers to the estimate of the universe total of the five secondary units in group #8. This would be repeated for the remaining sampled secondary units.

--- VARIANCE COMPONENTS FOR PRIMARY UNITS ---

WITHIN VARIANCE For each sampled primary unit, the contribution of the third-stage variation.

BETWEEN VARIANCE For each sampled primary unit, the contribution of the second-stage variation.

TOTAL VARIANCE For each sampled primary unit, the sum of WITHIN VARIANCE and BETWEEN VARIANCE. This value represents the total variation obtained by applying a RHC two-stage procedure to the sampled primary unit.

--- COMBINED VARIANCE COMPONENTS ---

STAGE 1	In the derivation of the standard error, the contribution of the first-stage (primary) units.
STAGES 2 AND 3	In the derivation of the standard error, the contribution of the second-stage (secondary) and third-stage units.
TOTAL VARIANCE	The sum of the values for STAGE 1 and STAGES 2 AND 3. The square root of this value is the STANDARD ERROR.

OVERALL SECTION: The results of the above information are then used in the final overall projection. The following information appears:

PRIMARY UNITS SAMPLED	The quantity of primary units selected in this sample.
PRIMARY UNITS NOT SAMPLED	The number of primary units in the population minus the number of primary units in the sample.
TOTAL PRIMARY UNITS	The total number of primary units in the population.
OVERALL POINT ESTIMATE	A single estimate for a universe value based on each primary unit point estimate multiplied by (A/B) where A is the size of the group containing the primary unit and B is the size of the primary unit. The products are summed over all primary units.
OVERALL STANDARD ERROR	A measurement of the standard deviation of the estimate for the population total. It is this value that determines the width of the corresponding confidence intervals.
CONFIDENCE LEVEL	This user's level of confidence (80%, 90%, or 95%) that the actual population total will fall within the corresponding confidence interval.
LOWER LIMIT	The lower bound of the confidence interval. It is based on subtracting the precision amount from the point estimate.
UPPER LIMIT	The upper bound of the confidence interval. It is based on adding the precision amount to the point estimate.

PRECISION AMOUNT A measurement of the closeness of the sample estimate and the corresponding population value. For a 90% confidence interval, the user would be 90% confident that the estimated population total (OVERALL POINT ESTIMATE) would be within this amount of the actual value. The precision amount is calculated by multiplying the standard error by the appropriate factor ("Z" value).

PRECISION PERCENT The result of dividing the precision amount by the point estimate.

Z-VALUE USED The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTRHC3.TXT, shown below. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/28/2004       RHC THREE-STAGE VARIABLE APPRAISAL            Time: 10:45
                          AUDIT/REVIEW: RHC 3-Stage

                     DATA FILE USED: C:\temp\DATARHC3.TXT
          PRIMARY/SECONDARY UNIVERSE FILE USED: C:\temp\PUSURHC3.TXT

                                                  *****D I F F E R E N C E*****
     **** SAMPLED UNITS ****        Third-stage  SAMPLE     SAMPLE      NONZERO
PRIMARY / SECONDARY IDENTIFICATION    UNIVERSE    SIZE       VALUE       COUNT
================================  ===========  ======  ============  ========
REGION3
    UNIV20                                 37       7         34.00         5
    UNIV38                                 74      15        116.00        12
    UNIV45                                 73      15        148.00        15
    UNIV10                                 60      12        103.00        11
    UNIV87                                 62      12         97.00        10
    UNIV82                                 30       6         28.00         4
    UNIV60                                 54      11         44.00         9
    UNIV69                                 39       8         84.00         8
    UNIV34                                 26       5         40.00         4
    UNIV54                                 57      11         92.00        10
  Total                                   512     102                      88
```

REGION7				
UNIV1	56	11	115.00	9
UNIV60	56	11	106.00	8
UNIV59	67	13	92.00	13
UNIV99	80	16	111.00	12
UNIV85	67	13	139.00	10
UNIV37	31	6	23.00	3
UNIV34	42	8	45.00	7
UNIV16	53	11	79.00	7
UNIV12	66	13	102.00	9
UNIV52	60	12	123.00	11
Total	578	114		89
REGION8				
UNIV19	34	7	34.00	6
UNIV104	77	15	108.00	13
UNIV66	49	10	46.00	9
UNIV110	38	8	72.00	7
UNIV83	70	14	80.00	10
UNIV14	48	10	80.00	9
UNIV78	27	5	52.00	5
UNIV105	27	5	54.00	5
UNIV12	65	13	65.00	11
UNIV112	75	15	84.00	12
Total	510	102		87
REGION9				
UNIV37	64	13	95.00	8
UNIV92	73	15	130.00	13
UNIV47	71	14	152.00	12
UNIV54	70	14	107.00	12
UNIV97	56	11	80.00	6
UNIV66	76	15	141.00	13
UNIV116	50	10	76.00	6
UNIV29	33	7	63.00	6
UNIV18	26	5	20.00	2
UNIV107	55	11	125.00	10
Total	574	115		88
TOTALS	2,174	433	3,385.00	352

**** SAMPLED UNITS **** PRIMARY / SECONDARY IDENTIFICATION	SAMPLE MEAN	SIZES RATIO	POINT ESTIMATE
REGION3			
UNIV20	4.86	12.1667	2,187
UNIV38	7.73	6.3636	3,642
UNIV45	9.87	7.4545	5,369
UNIV10	8.58	9.0000	4,635
UNIV87	8.08	5.9000	2,957
UNIV82	4.67	13.6000	1,904
UNIV60	4.00	8.3333	1,800
UNIV69	10.50	10.8571	4,446
UNIV34	8.00	15.0000	3,120
UNIV54	8.36	8.4444	4,026
TOTAL			34,085

REGION7			
UNIV1	10.45	8.9000	5,211
UNIV60	9.64	9.6000	5,181
UNIV59	7.08	7.2308	3,428
UNIV99	6.94	6.5000	3,608
UNIV85	10.69	7.1538	5,125
UNIV37	3.83	17.1667	2,040
UNIV34	5.63	13.2500	3,130
UNIV16	7.18	8.3000	3,159
UNIV12	7.85	6.9231	3,585
UNIV52	10.25	10.4545	6,430
TOTAL			40,896
REGION8			
UNIV19	4.86	17.1250	2,828
UNIV104	7.20	7.6250	4,227
UNIV66	4.60	11.0909	2,500
UNIV110	9.00	15.2222	5,206
UNIV83	5.71	8.0667	3,227
UNIV14	8.00	14.7000	5,645
UNIV78	10.40	17.4286	4,894
UNIV105	10.80	24.0000	6,998
UNIV12	5.00	8.7143	2,832
UNIV112	5.60	7.8750	3,308
TOTAL			41,665
REGION9			
UNIV37	7.31	10.5000	4,911
UNIV92	8.67	8.3333	5,272
UNIV47	10.86	8.6667	6,681
UNIV54	7.64	8.7333	4,672
UNIV97	7.27	11.5000	4,684
UNIV66	9.40	7.6250	5,447
UNIV116	7.60	14.0000	5,320
UNIV29	9.00	16.0000	4,752
UNIV18	4.00	18.8571	1,961
UNIV107	11.36	11.5455	7,216
TOTAL			50,916

--- VARIANCE COMPONENTS FOR PRIMARY UNITS ---

**** SAMPLED UNITS **** PRIMARY UNIT IDENTIFICATION	WITHIN VARIANCE	BETWEEN VARIANCE	TOTAL VARIANCE
REGION3	434,942	9,869,853	10,304,795
REGION7	892,072	14,293,056	15,185,128
REGION8	499,759	14,612,210	15,111,969
REGION9	1,134,688	20,534,696	21,669,384

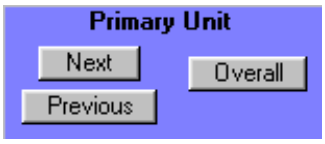
```
                    --- COMBINED VARIANCE COMPONENTS ---

        STAGE 1              STAGES 2 AND 3         TOTAL VARIANCE
=======================  =======================  =======================

          1,148,658,590             170,994,089           1,319,652,679


                    --- SUMMARY OF APPRAISAL RESULTS ---
PRIMARY UNITS SAMPLED                                        4
PRIMARY UNITS NOT SAMPLED                                    8
TOTAL PRIMARY UNITS                                         12
OVERALL POINT ESTIMATE                                 476,331
OVERALL STANDARD ERROR                                  36,327

                                        CONFIDENCE LIMITS
                                       80% CONFIDENCE LEVEL
           LOWER LIMIT                                429,776
           UPPER LIMIT                                522,886
           PRECISION AMOUNT                            46,555
           PRECISION PERCENT                            9.77%
           Z-VALUE USED                       1.281551565545

                                       90% CONFIDENCE LEVEL
           LOWER LIMIT                                416,579
           UPPER LIMIT                                536,084
           PRECISION AMOUNT                            59,753
           PRECISION PERCENT                           12.54%
           Z-VALUE USED                       1.644853626951

                                       95% CONFIDENCE LEVEL
           LOWER LIMIT                                405,132
           UPPER LIMIT                                547,531
           PRECISION AMOUNT                            71,200
           PRECISION PERCENT                           14.95%
           Z-VALUE USED                       1.959963984540
```

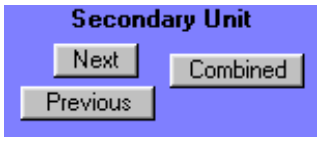
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

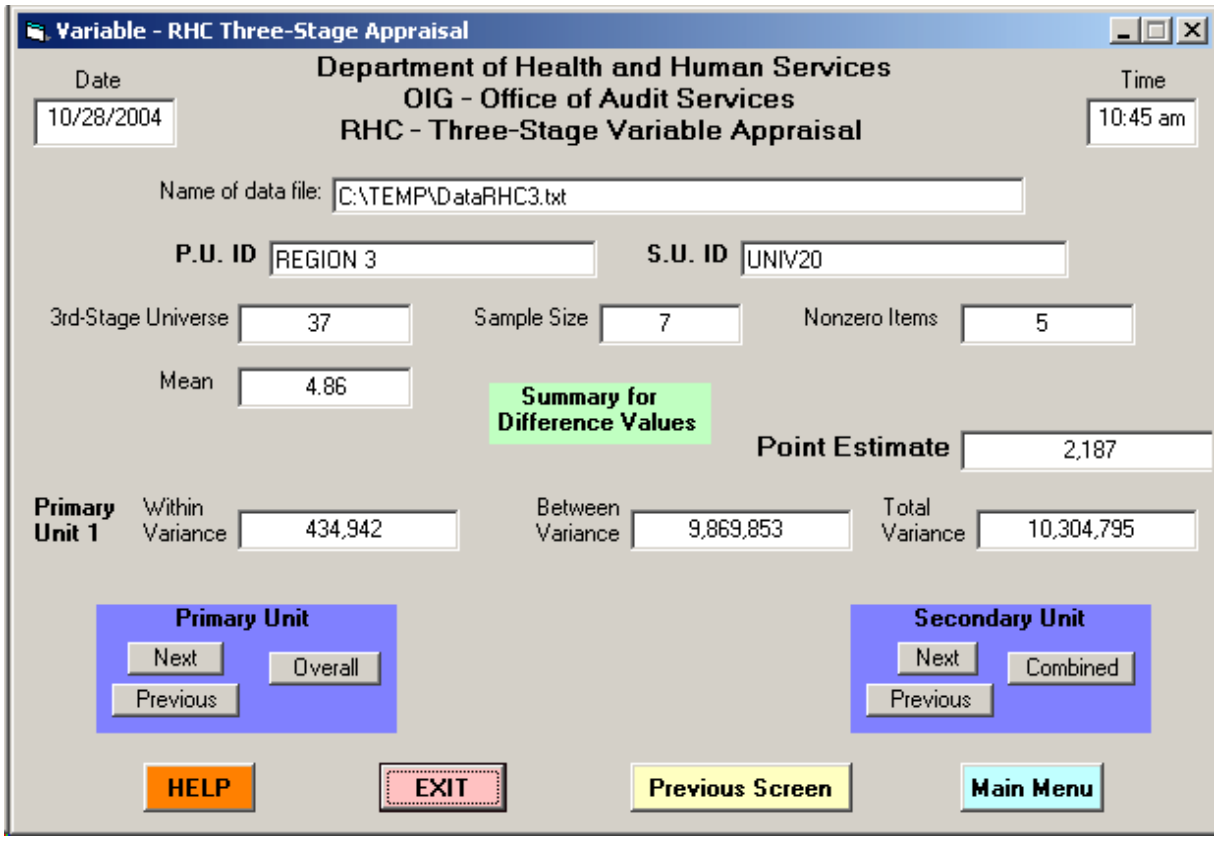
Output to Screen

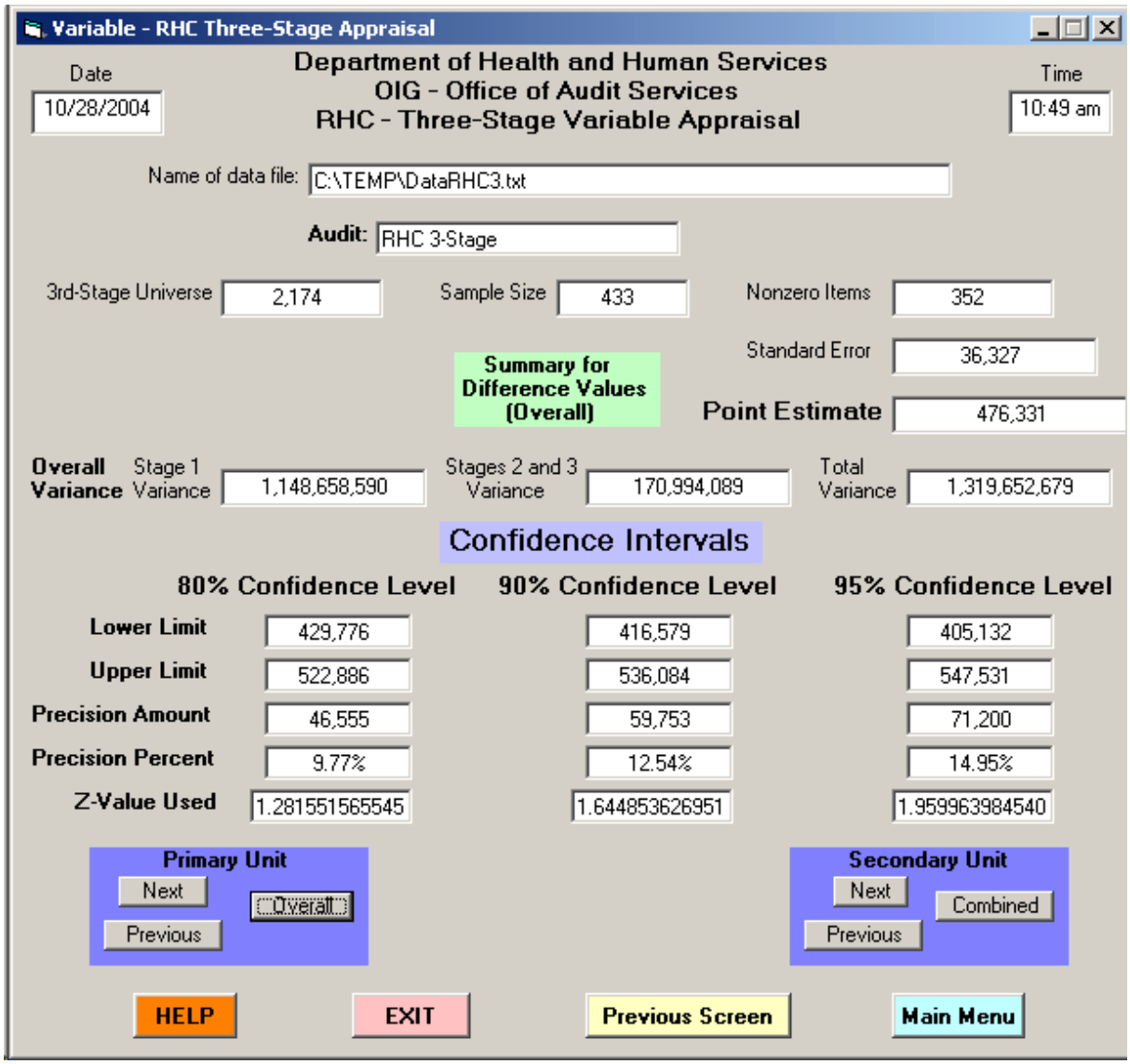
The program always concludes with a summary on the screen, even if the output option selected is a text file or printer. The screen below is the summary of the difference values for the first primary unit in this illustration. If the user created a sample data file containing two values per sample item (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 form. For this illustration, the **Additional Summary Info** button is not visible since only the difference values were used in the data file.

To view summary information for the next sampled secondary unit within the selected primary unit, click on **Next** in the bottom **Secondary Unit** box. By clicking on **Next** and **Previous**, the user can navigate through the secondary units for the selected primary unit. By clicking on **Combined**, the screen will contain the summary for the selected primary unit. Similarly, by clicking on **Next** and **Previous** in the **Primary Unit** box, the user can navigate through the sampled primary units. For each selected primary unit, the user can then switch to the **Secondary Unit** box and examine the results for each secondary unit within this primary unit. Finally, by clicking on **Overall** in the **Primary Unit** box, the summary appraisal results for the entire analysis are shown. These output screens are shown on the next two pages.









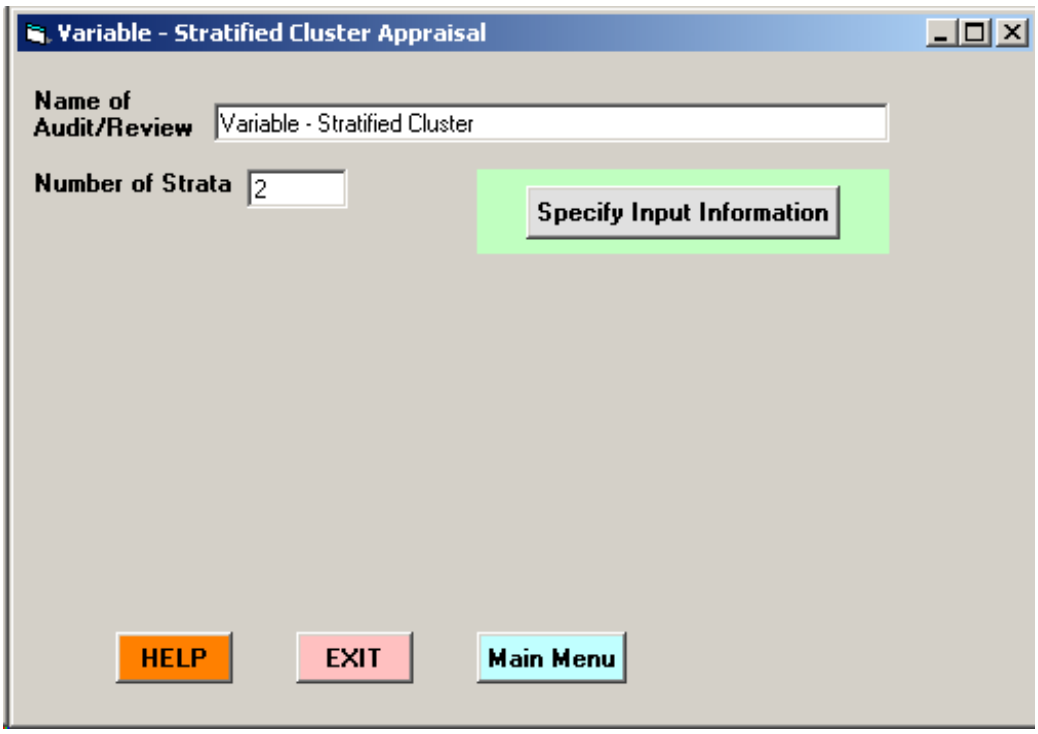
NOTE Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

STRATIFIED CLUSTER

Purpose

This program performs a stratified cluster appraisal of a previously created data file based on information gathered from all items in sampled clusters. The user would have initially stratified the clusters (e.g., universities) into two or more categories (e.g., public and private universities). Within each stratum, the user would randomly select clusters. Then all items (e.g., grants) within the cluster would be reviewed.

Input Screen



Name of audit/review

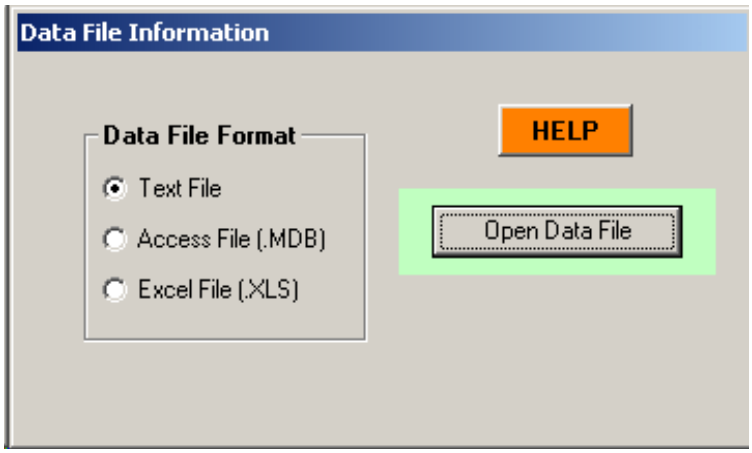
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

Number of strata

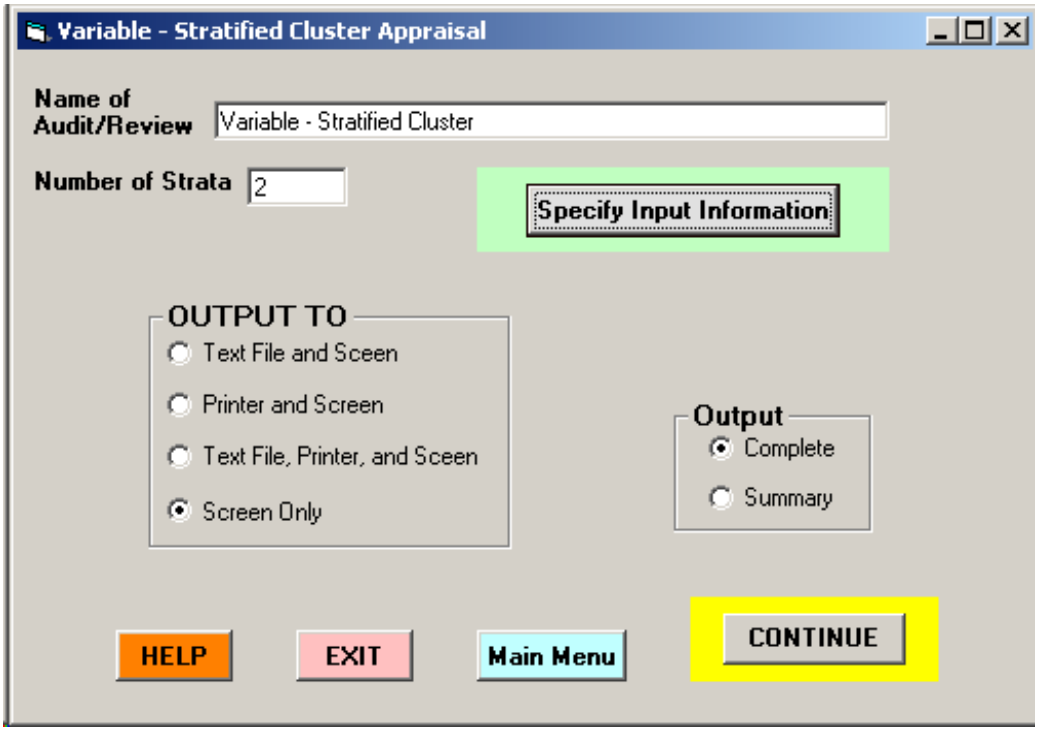
After entering the name of the audit/review, the user must enter the number of strata. The maximum number of strata is 100.

Specify input information

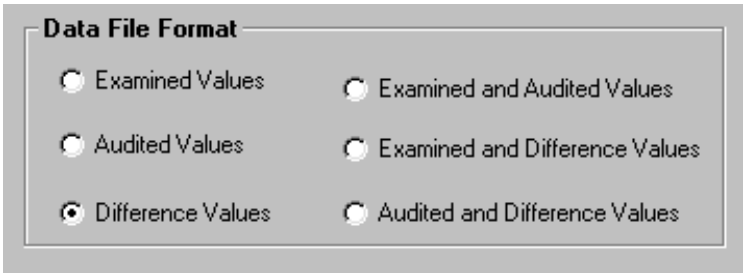
After entering the above information, click on **Specify Input Information**. The following screen will appear. The input file format can be a text file, a table within an Access database, or an Excel spreadsheet. After selecting the desired format, click on **Open Data File**. Using the standard Windows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



When the file has been opened, the full input screen (shown next) will appear.



Format of Input File



Prior to executing this program the user must create a sample data file that contains certain identifying data and sample information for each cluster selected. The identifying data are descriptive identifiers of the stratum and each cluster (maximum of 25 characters for each identifier). For each stratum the user enters the number of clusters (e.g., universities) that makes up the universe for the stratum. For each cluster sampled the user enters the number of items (e.g., grants) in the cluster and the total value for all items in the cluster of the quantitative characteristic being measured (e.g., unauthorized travel costs).

For each stratum the user needs to use the following format:

PRIVATE UNIVERSITIES 415 25

Explanation:

PRIVATE - UNIVERSITIES This is a description of the clusters in the stratum. It must be least one character in length. It may be longer than 25 characters in length; however, only the first 25 characters will be stored for output.

415 - This is the number of clusters in the universe for the stratum.

25 - This is the number of clusters in the sample for the stratum.

The data values may contain commas and dollar signs ($). The program assumes one or more spaces or tabs as the only delimiters between pieces of data.

Each stratum data line is followed by one or more lines containing data for each cluster sampled. The format for each cluster is as follows:

UNIV1 8 96

Explanation:

UNIV1 - This is a description of the sampled cluster. It must be at least one character in length. It may be longer than 25 characters in length; however, only the first 25 characters will be stored for output.

8 - This is the number of items reviewed within the cluster. The quantity must be separated from the cluster description by one or more spaces. Commas are allowed for this value.

96 - This is the total value for the quantitative characteristic being measured for the items in the cluster. The value must be separated by one or more spaces from the number of items reviewed in the cluster. The value may contain commas, a decimal point, or a dollar sign. If the value is negative, it must be preceded by a minus sign (-).

Input From a Text File

The sample data 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).

As an illustration, consider two strata: Stratum 1: state-supported universities (N_1 = 415 universities in the population) and Stratum 2: private universities (N_2 = 168 universities in the population). Within each stratum, a single-stage cluster sample was obtained with n_1 = 25 universities selected from Stratum 1 and n_2 = 10 universities from Stratum 2. For each of the sampled universities, all health-related grants would be audited to determine the amount of costs improperly charged to these grants. The sample data file (C:\TEMP\DATASTRCLUS.TXT) is shown below.

```
STATE UNIVERSITIES  415  25
UNIV1     8   96
UNIV2    12  121
UNIV3     4   42
UNIV4     5   65
UNIV5     6   52
UNIV6     6   40
UNIV7     7   75
UNIV8     5   65
UNIV9     8   45
UNIV10    3   50
UNIV11    2   85
UNIV12    6   43
UNIV13    5   54
UNIV14   10   49
UNIV15    9   53
UNIV16    3   50
UNIV17    6   32
UNIV18    5   22
UNIV19    5   45
UNIV20    4    37
UNIV21    6    51
UNIV22    8    30
UNIV23    7    39
UNIV24    3    47
UNIV25    8    41
PRIVATE UNIVERSITIES 168 10
UNIV1     2    18
UNIV2     5    52
UNIV3     7    68
UNIV4     4    36
UNIV5     3    45
UNIV6     8    96
UNIV7     6    64
UNIV8    10   115
UNIV9     3    41
UNIV10    1    12
```

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

Input From an Access Database

The sample data 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 on the standard Windows "Open" file screen.

For this illustration, the data will be stored in table DATASTRCLUS within Access database C:\TEMP\VARIABLE.MDB. The first seven rows of Access table DATASTRCLUS are shown below. The final 15 rows immediately follow.

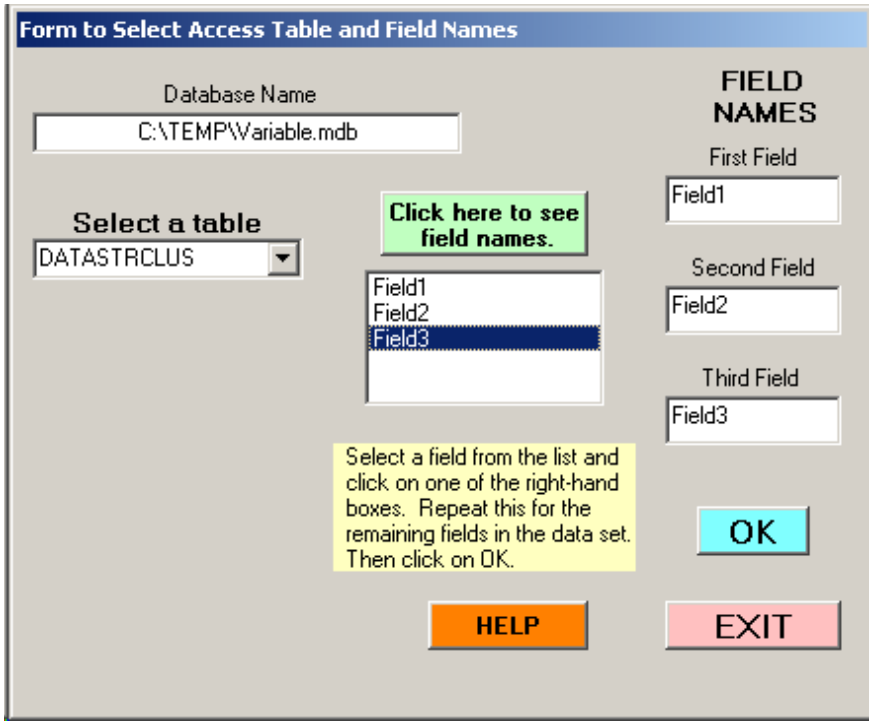
DATASTRCLUS : Table

Field1	Field2	Field3
STATE UNIVERSITIES	415	25
UNIV1	8	96
UNIV2	12	121
UNIV3	4	42
UNIV4	5	65
UNIV5	6	52
UNIV6	6	40
UNIV22	8	30
UNIV23	7	39
UNIV24	3	47
UNIV25	8	41
PRIVATE UNIVERSITIES	168	10
UNIV1	2	18
UNIV2	5	52
UNIV3	7	68
UNIV4	4	36
UNIV5	3	45
UNIV6	8	96
UNIV7	6	64
UNIV8	10	115
UNIV9	3	41
UNIV10	1	12

Once the data file is open, the program returns to the input screen. Click on **CONTINUE** to resume processing. 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 (DATASTRCLUS for this illustration) and clicking on **Click here to see field names**, the field names will appear.

To select the field names, click on the field name for the field containing the primary and secondary unit IDs ("Field1" in this illustration) and then click on the top right-hand box. The field name will then appear in this box. Repeat this procedure for the remaining two field names in this table ("Field2" and "Field3" in this illustration). When all the field names have been specified, click on **OK**. The program will continue processing.



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

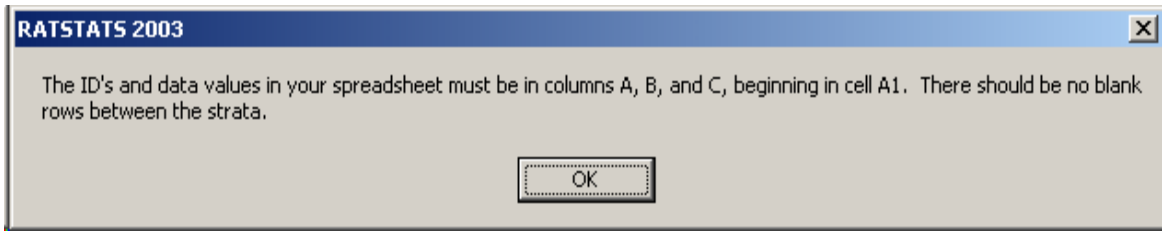
Input From an Excel Spreadsheet

The input file can be stored in an Excel spreadsheet containing three columns. 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 DATASTRCLUS.XLS will be used. The first seven rows of this spreadsheet are shown below. The final 15 rows immediately follow.

	A	B	C
1	STATE UNIVERSITIES	415	25
2	UNIV1	8	96
3	UNIV2	12	121
4	UNIV3	4	42
5	UNIV4	5	65
6	UNIV5	6	52
7	UNIV6	6	40
23	UNIV22	8	30
24	UNIV23	7	39
25	UNIV24	3	47
26	UNIV25	8	41
27	PRIVATE UNIVERSITIES	168	10
28	UNIV1	2	18
29	UNIV2	5	52
30	UNIV3	7	68
31	UNIV4	4	36
32	UNIV5	3	45
33	UNIV6	8	96
34	UNIV7	6	64
35	UNIV8	10	115
36	UNIV9	3	41
37	UNIV10	1	12

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

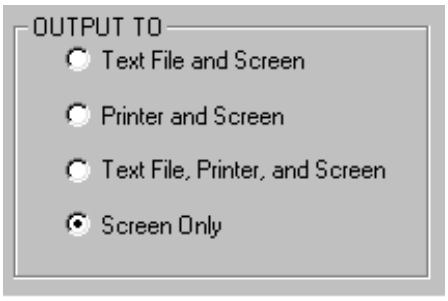
When using the Excel option, the following screen will be displayed.



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 a text file and screen; a printer and screen; a text file, printer, and screen; or 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.

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**.

Program Output

Whether the user selects the complete or summary option for output, the following information will be displayed for each stratum:

STRATUM IDENTIFICATION	The descriptions entered in the sample data file.
SAMPLE UNIVERSE	The universe size of clusters for each stratum. The numbers were obtained from the data file.
SAMPLE SIZE	The count of clusters in the data file for each stratum. It represents the number of clusters reviewed in each stratum.

If the user selects the complete option, the following information will be displayed:

CLUSTER IDENTIFICATION	The descriptions entered in the data file.
SAMPLE UNIVERSE	The universe size for each cluster in a stratum. The numbers were obtained from the data file.
SAMPLE SIZE	The number of items actually reviewed. Since this is a cluster sample, the universe and sample sizes are the same within each cluster. The number of items may vary from cluster to cluster, but all items in all sampled clusters are to be reviewed.
SAMPLED VALUE	The summation of values entered in the data file for the quantitative characteristic being measured. It represents the summation for each cluster.

The final two pieces of information for each stratum (sampled value and point estimate) will appear on different lines depending on the output option selected (i.e., complete or summary). If a complete option is selected, then after displaying the information for each sampled cluster in a stratum, a stratum totals line will appear containing these two values. If, however, the summary option was selected, then these two values will appear on the same line as the stratum identification.

SAMPLED VALUE	The summation of values entered in the data file for the quantitative characteristic being measured. In this line it represents the summation for the stratum.
POINT ESTIMATE	A single estimate for the stratum of the universe value of the characteristic being measured. It is obtained by dividing the summation of stratum sampled values by the number of clusters sampled and multiplying the result by the number of clusters in the stratum universe.

Whether the complete or summary option is selected, the following output will be displayed after all the information for the strata are shown:

STRATA TOTALS	The total number of clusters in the universe and the actual number of clusters sampled.
CLUSTER UNIT TOTALS	Information for all clusters sampled including the total number of items reviewed and the value of those items for the characteristic being measured.
OVERALL POINT ESTIMATE	A single estimate of the overall value of the characteristic being measured. It is calculated by the summation of the point estimates for all strata.
OVERALL STANDARD ERROR	A measurement of the variance of the overall point estimate. It is this value that is used in determining the width of the confidence intervals.
CONFIDENCE LEVEL	The user's level of confidence that the actual value of the measured characteristic will fall within the range from the lower to the upper limits (confidence interval).
LOWER LIMIT	The lower bound of the confidence interval. It is calculated by subtracting the precision amount from the point estimate.
UPPER LIMIT	The upper bound of the confidence interval. It is calculated 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 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 The result of dividing the precision amount by the point estimate.

Z-VALUE USED The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

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

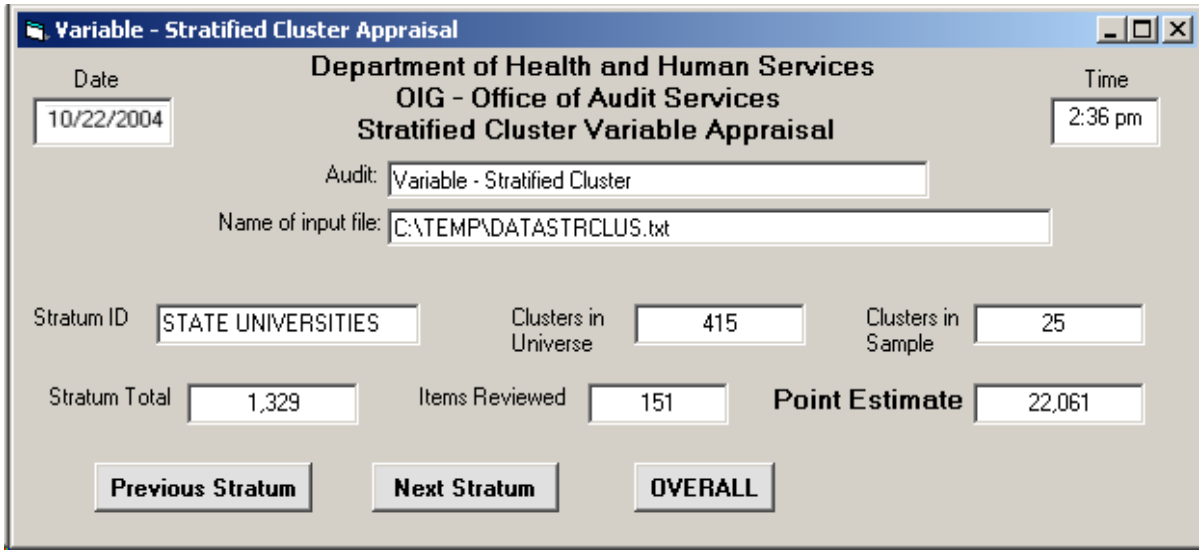
```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004      STRATIFIED CLUSTER VARIABLE APPRAISAL        Time: 14:36
                 AUDIT/REVIEW: Variable - Stratified Cluster

                   DATA FILE USED: C:\TEMP\DATASTRCLUS.TXT

STRATUM IDENTIFICATION          SAMPLE    SAMPLE
  CLUSTER IDENTIFICATION       UNIVERSE    SIZE   SAMPLED VALUE  POINT ESTIMATE
===========================  ===========  ======  =============  ==============
STATE UNIVERSITIES                   415      25
  UNIV1                                8       8          96.00
  UNIV2                               12      12         121.00
  UNIV3                                4       4          42.00
  UNIV4                                5       5          65.00
  UNIV5                                6       6          52.00
  UNIV6                                6       6          40.00
  UNIV7                                7       7          75.00
  UNIV8                                5       5          65.00
  UNIV9                                8       8          45.00
  UNIV10                               3       3          50.00
  UNIV11                               2       2          85.00
  UNIV12                               6       6          43.00
  UNIV13                               5       5          54.00
  UNIV14                              10      10          49.00
  UNIV15                               9       9          53.00
  UNIV16                               3       3          50.00
  UNIV17                               6       6          32.00
  UNIV18                               5       5          22.00
  UNIV19                               5       5          45.00
  UNIV20                               4       4          37.00
  UNIV21                               6       6          51.00
  UNIV22                               8       8          30.00
  UNIV23                               7       7          39.00
```

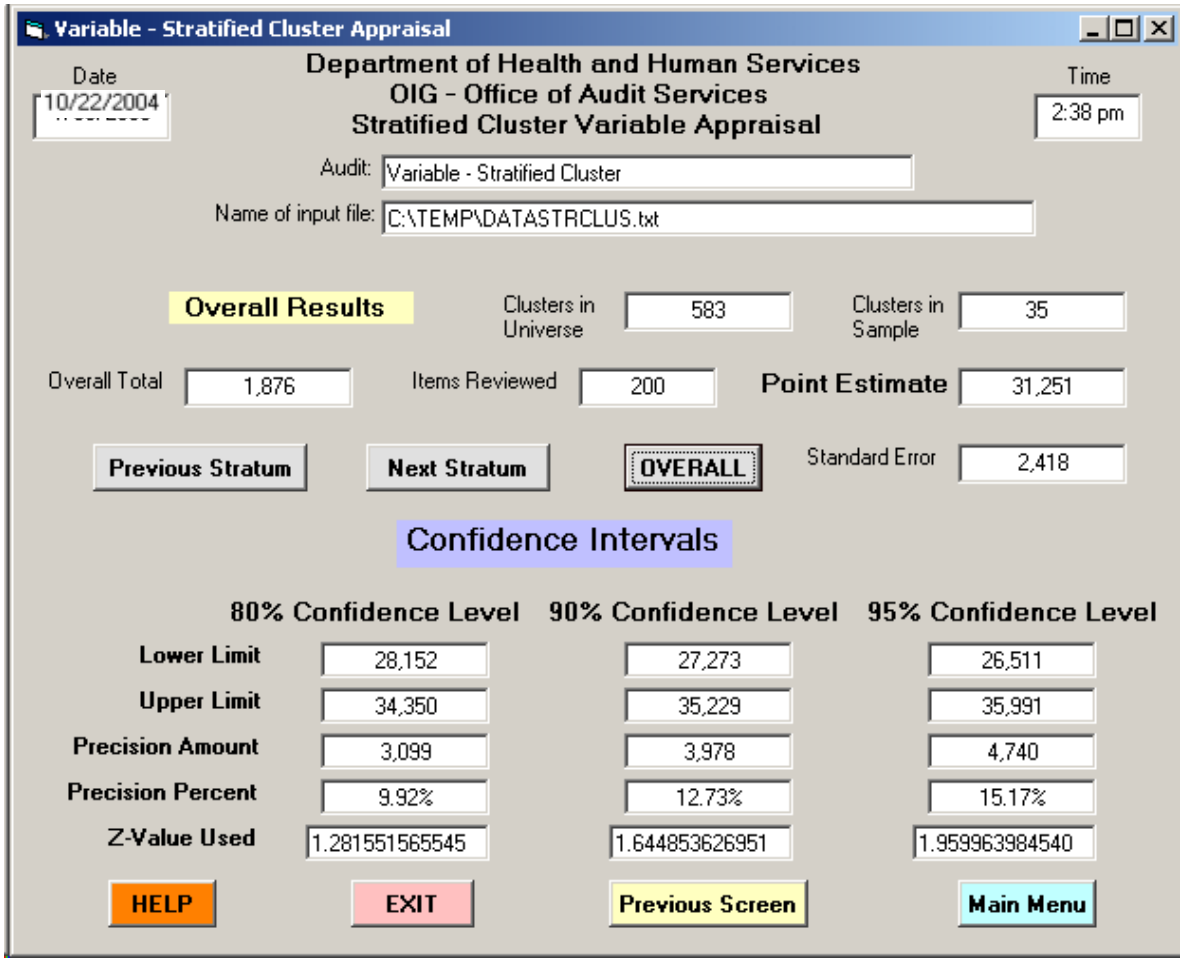
```
  UNIV24                              3       3          47.00
  UNIV25                              8       8          41.00
      STRATUM TOTALS                151     151       1,329.00          22,061

     10/22/2004 PRIVATE UNIVERSITIES                  168      10
                UNIV1                                   2       2          18.00
  UNIV2                               5       5          52.00
  UNIV3                               7       7          68.00
  UNIV4                               4       4          36.00
  UNIV5                               3       3          45.00
  UNIV6                               8       8          96.00
  UNIV7                               6       6          64.00
  UNIV8                              10      10         115.00
  UNIV9                               3       3          41.00
  UNIV10                              1       1          12.00
      STRATUM TOTALS                 49      49         547.00           9,190


STRATUM IDENTIFICATION          SAMPLE    SAMPLE
  CLUSTER IDENTIFICATION       UNIVERSE    SIZE   SAMPLED VALUE  POINT ESTIMATE
===========================  ===========  ======  =============  ==============

STRATA TOTALS                       583      35
CLUSTER UNIT TOTALS                 200     200       1,876.00
OVERALL POINT ESTIMATE                                                  31,251
OVERALL STANDARD ERROR                                                   2,418

CONFIDENCE LEVEL      ---80 PERCENT--      ---90 PERCENT--      ---95 PERCENT--
LOWER LIMIT                   28,152               27,273               26,511
UPPER LIMIT                   34,350               35,229               35,991
PRECISION AMOUNT               3,099                3,978                4,740
PRECISION PERCENT              9.92%               12.73%               15.17%
Z-VALUE USED          1.281551565545       1.644853626951       1.959963984540
```

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

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. 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**. The resulting summary screen immediately follows the summary for the first stratum.


Variable - Stratified Cluster Appraisal
Date
10/22/2004
Department of Health and Human Services
OIG - Office of Audit Services
Stratified Cluster Variable Appraisal
Time
2:36 pm
Audit: Variable - Stratified Cluster
Name of input file: C:\TEMP\DATASTRCLUS.txt
Stratum ID STATE UNIVERSITIES
Clusters in Universe 415
Clusters in Sample 25
Stratum Total 1,329
Items Reviewed 151
Point Estimate 22,061
Previous Stratum
Next Stratum
OVERALL

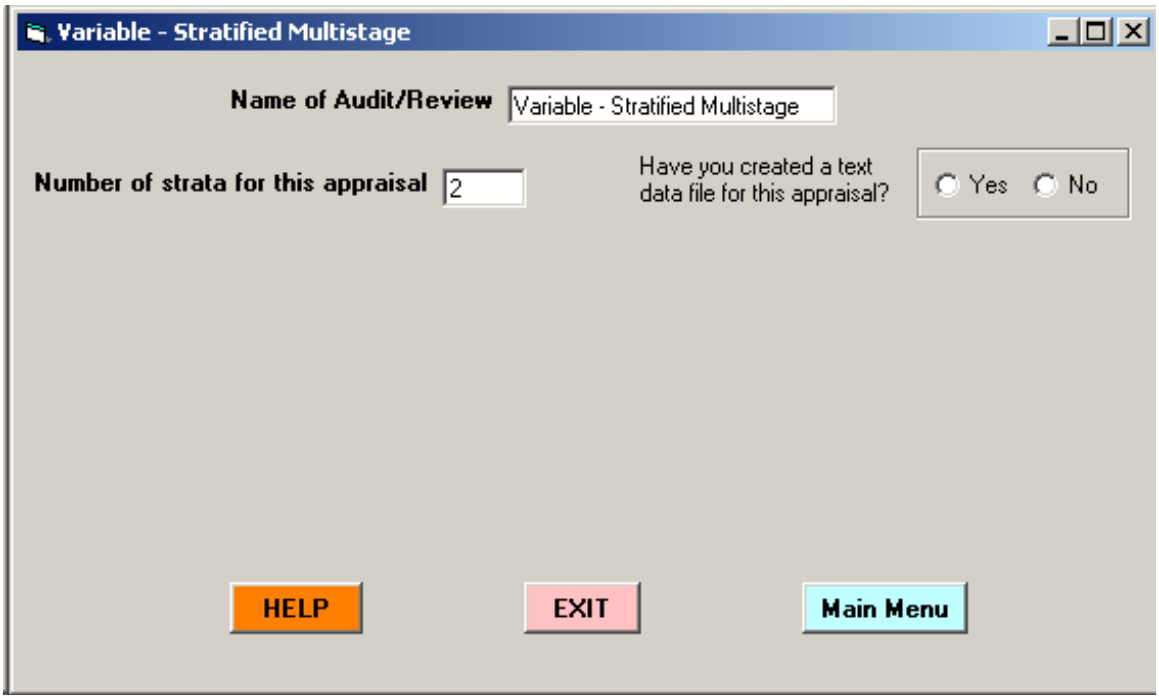

Variable - Stratified Cluster Appraisal
Date
10/22/2004
Department of Health and Human Services
OIG - Office of Audit Services
Stratified Cluster Variable Appraisal
Time
2:38 pm
Audit: Variable - Stratified Cluster
Name of input file: C:\TEMP\DATASTRCLUS.txt
Overall Results
Clusters in Universe 583
Clusters in Sample 35
Overall Total 1,876
Items Reviewed 200
Point Estimate 31,251
Previous Stratum
Next Stratum
OVERALL
Standard Error 2,418
Confidence Intervals
80% Confidence Level
90% Confidence Level
95% Confidence Level
Lower Limit 28,152 27,273 26,511
Upper Limit 34,350 35,229 35,991
Precision Amount 3,099 3,978 4,740
Precision Percent 9.92% 12.73% 15.17%
Z-Value Used 1.281551565545 1.644853626951 1.959963984540
HELP
EXIT
Previous Screen
Main Menu

STRATIFIED MULTISTAGE

Purpose

This program performs a stratified multistage appraisal based on information gathered from prior multistage appraisals. The user would have initially stratified the clusters (e.g., universities) into two or more categories (e.g., public and private universities). Within each stratum, the user would select a multistage sample. The results of the samples would be appraised using a multistage appraisal program. The point estimate and standard error from each of these appraisals could be placed in a data file or entered interactively by the user. **NOTE**: Use the point estimates and standard errors for the totals, not the point estimates and standard errors for the means.

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 for this appraisal

Enter the number of strata that will be appraised. This number must be between 2 and 500.

Have you created a text data file for this appraisal?

Prior to executing this program the user could create a data file that contains the point estimate and the standard error for each of the strata. The data file containing the above information must be stored in a text file format. There are several ways the user may create this file. The easiest approach would be with a text editor or word processing package (e.g., Word) or a print file created with a spreadsheet package (e.g., Excel).

For each stratum the user needs to use the following format:

24,077 1,277

Explanation:

24,077 - The point estimate of the total for one stratum. The program will allow commas to be included in the value.

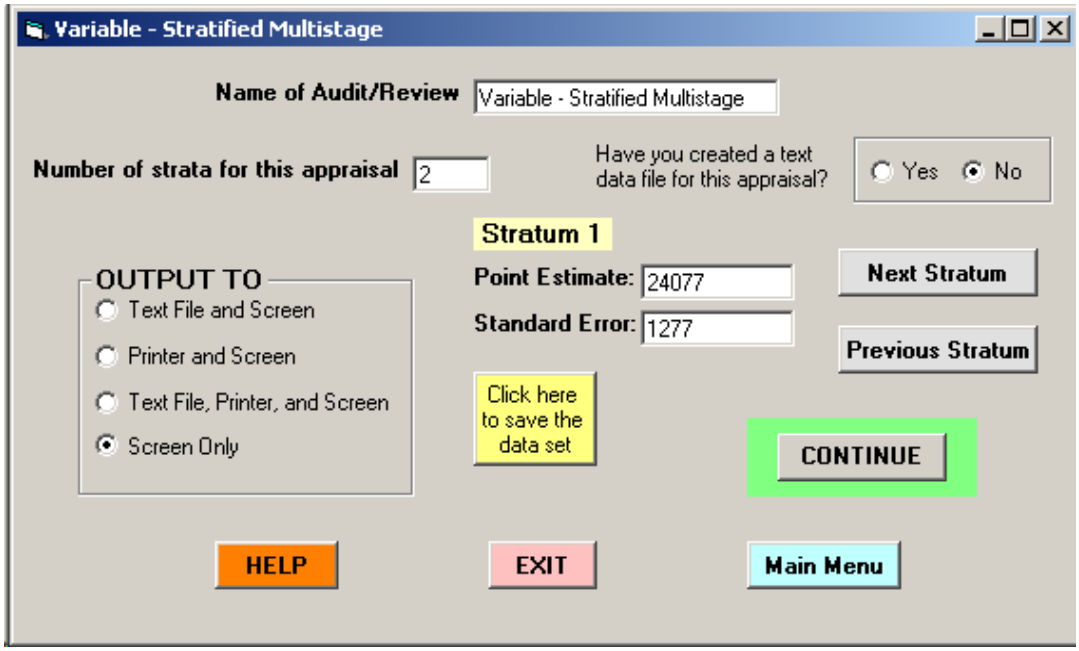
1,277 - The standard error of the total estimate for one stratum. The program will allow commas to be included in the value.

The results for each stratum should be on a separate line and the values separated by one or more spaces or tabs.

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, then click on **Open**. If a file name is entered, the values obtained from the file will be displayed on the screen for review by the user.

If **No** is selected at this query the following screen will appear:

Input of Data Values



Entering point estimates

For the stratum identified on the screen, the user may enter a point estimate of the stratum total.

Entering standard errors

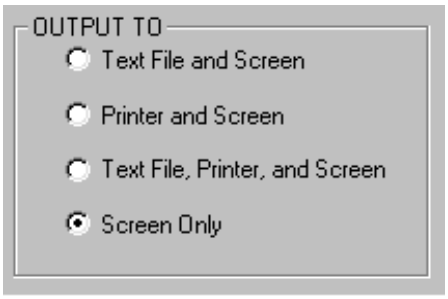
For the stratum identified on the screen, the user may enter a standard error for the estimate of the stratum total.

Editing the values

By clicking on **Next Stratum** and **Previous Stratum**, the user can navigate through the strata to observe, enter, or modify the point estimates and standard errors.

At any point before clicking on the **CONTINUE** button, the user may change the point estimate or standard error for any of the strata. When all edits are complete, click on **Click here to save the data set**. The standard Windows "Save As" screen will appear. Type the output file name alongside the **File name** box and click on **Save**.

Output Options



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or 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.

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**.

Program Output

The output will first display the data entered by the user for each of the strata. Then the output will display the results of the appraisal. The following will be displayed in the results portion:

POINT ESTIMATE	A single estimate for all strata of the universe value of the characteristic being measured. It is obtained by the summation of strata point estimates.
STANDARD ERROR	A measurement of the variation of the point estimate of the universe total with respect to all possible point estimates for this universe and these sample sizes.
CONFIDENCE LEVEL	The user's level of confidence that the actual value of the measured characteristic will fall within the range from the lower to upper limits (confidence interval).

LOWER LIMIT	The lower bound of the confidence interval. It is calculated by subtracting the precision amount from the point estimate.
UPPER LIMIT	The upper bound of the confidence interval. It is calculated by adding the precision amount to the point estimate.
PRECISION AMOUNT	A measurement of the closeness of the sample estimate and the corresponding population value. For a 90% confidence interval, the user would be 90% confident that the estimated population total (OVERALL POINT ESTIMATE) would be within this amount of the actual value. The precision amount is calculated by multiplying the standard error by the appropriate factor ("Z" value).
PRECISION PERCENT	The result of dividing the precision amount by the point estimate.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

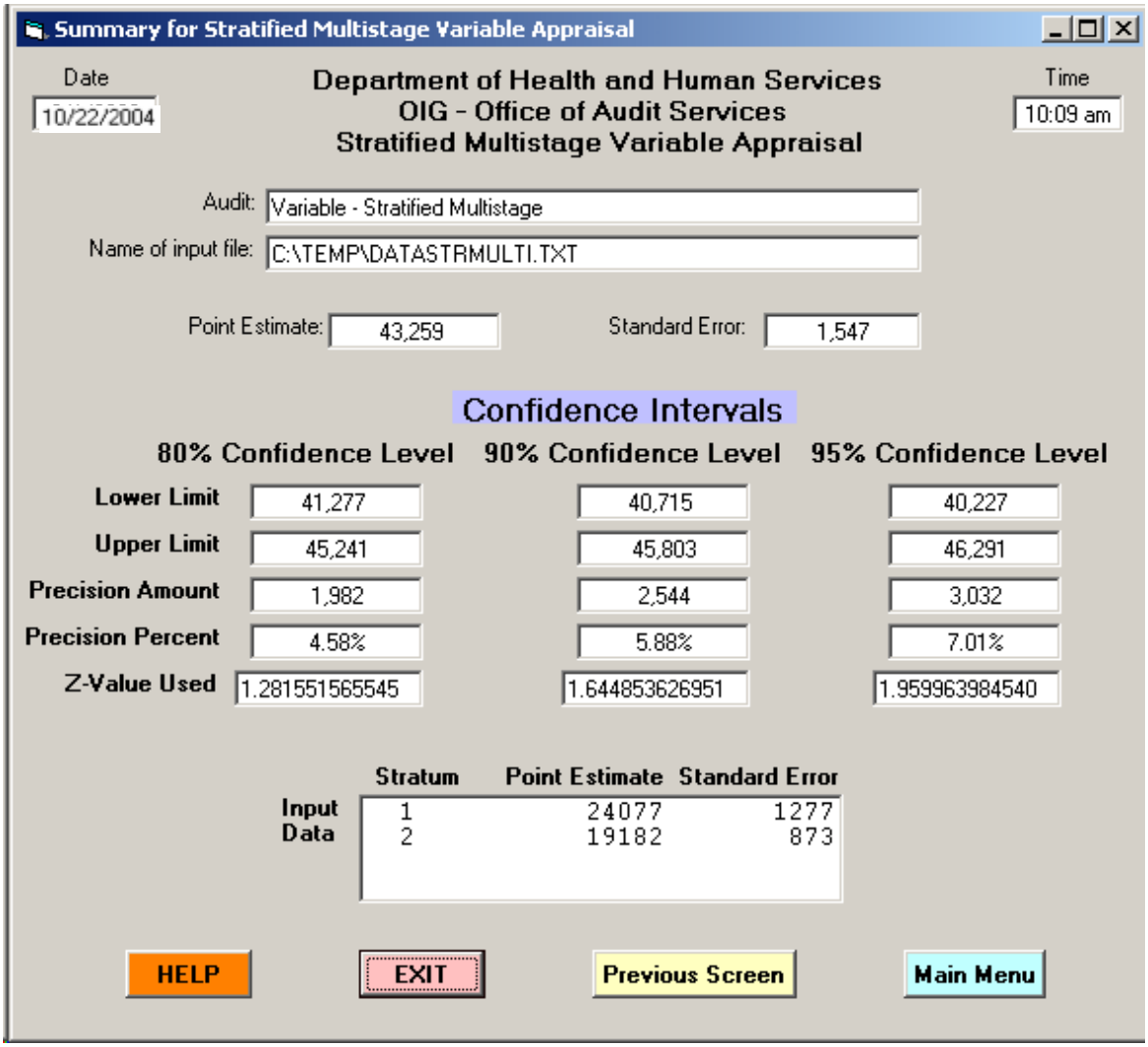
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTSTRMULTI.TXT, shown below. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004   STRATIFIED MULTISTAGE VARIABLE APPRAISAL        Time: 10:09
                AUDIT/REVIEW: Variable - Stratified Multistage
                    DATA FILE: C:\TEMP\DATASTRMULTI.TXT

             THE ESTIMATORS ARE BASED ON THE FOLLOWING ENTRIES:
  STRATUM        POINT ESTIMATE     STANDARD ERROR
      1                24,077              1,277
      2                19,182                873


= = = = = = = = = = = = = = = = = = RESULTS = = = = = = = = = = = = = = = = = =

                 POINT ESTIMATE              STANDARD ERROR

                        43,259                       1,547

CONFIDENCE LEVEL      ---80 PERCENT--      ---90 PERCENT--      ---95 PERCENT--
LOWER LIMIT                  41,277               40,715               40,227
UPPER LIMIT                  45,241               45,803               46,291
PRECISION AMOUNT              1,982                2,544                3,032
PRECISION PERCENT             4.58%                5.88%                7.01%
Z-VALUE USED          1.281551565545       1.644853626951       1.959963984540
```

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

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 summary screen for this illustration is shown on the next page. The input values are contained in the box at the bottom of the screen.



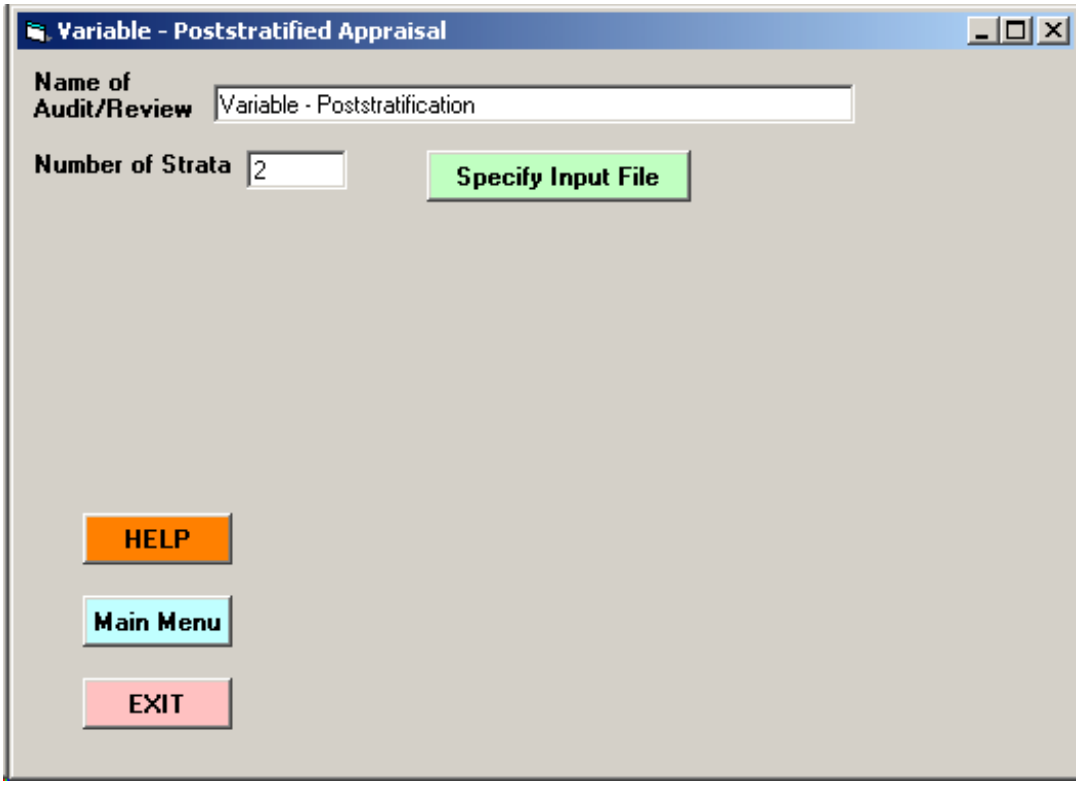
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

POSTSTRATIFICATION

Purpose

While in many cases the user would like to develop a stratified sampling plan, it may not be feasible to do so before actually drawing the sample, or the user may not recognize the need to stratify until after the sample has been drawn and the items evaluated. In such situations post stratification may be used. Such a stratification methodology may only be used if the sizes of the strata populations are known and each stratum sample is of sufficient size. This method, however, is less efficient statistically than a stratified sample.

Input Screen



Name of audit/review

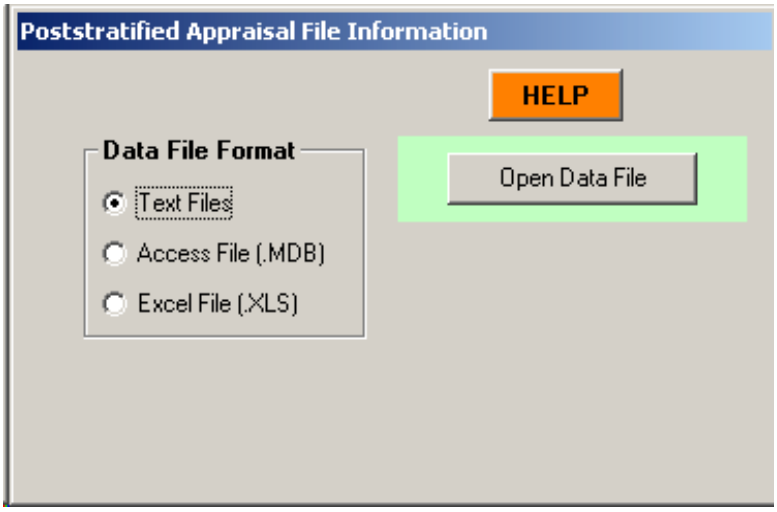
This program allows the user to enter a brief description of the audit or purpose of the evaluation. The description, which appears at the top of the output, is limited to 40 characters in length and may include commas and spaces.

Number of strata

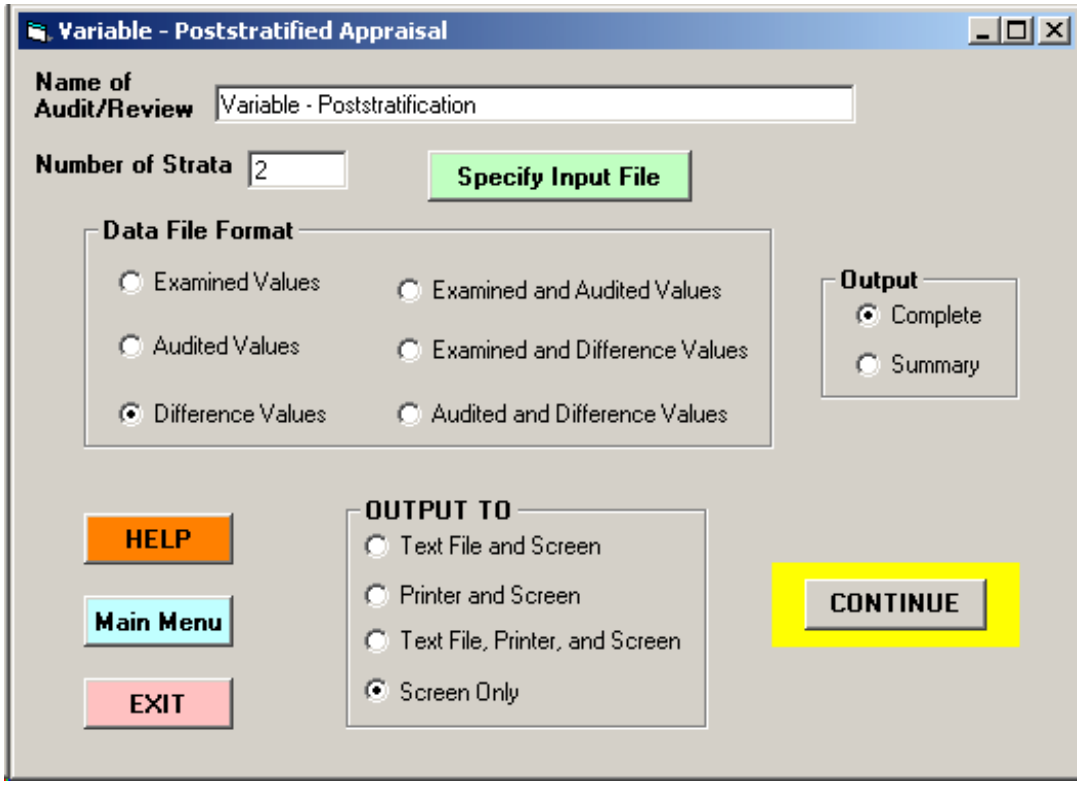
After entering the name of the audit/review, the user must enter the number of strata.

Specify input file

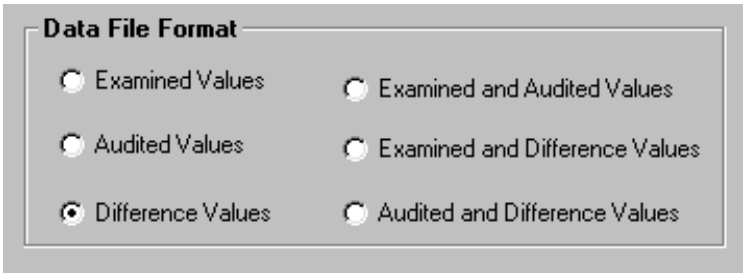
After entering the above information, click on **Specify Input File**. The following screen will appear. The input file(s) format can consist of two text files, two tables within an Access database, or two spreadsheets within the same Excel file. After selecting the desired format, click on **Open Data File**. Using the standard Windows "Open" file screen, locate the input file and double-click on it.



When all files have been opened, the full input screen (shown next) will appear.



Format of Input File



Prior to executing this program, the user must create a sample 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 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.

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 must enter the data for each sampling unit on a separate line. For ease of editing, enough spaces or tabs 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 or tabs as the only delimiters between pieces of data.

Input From Text Files

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).

Sample Data File

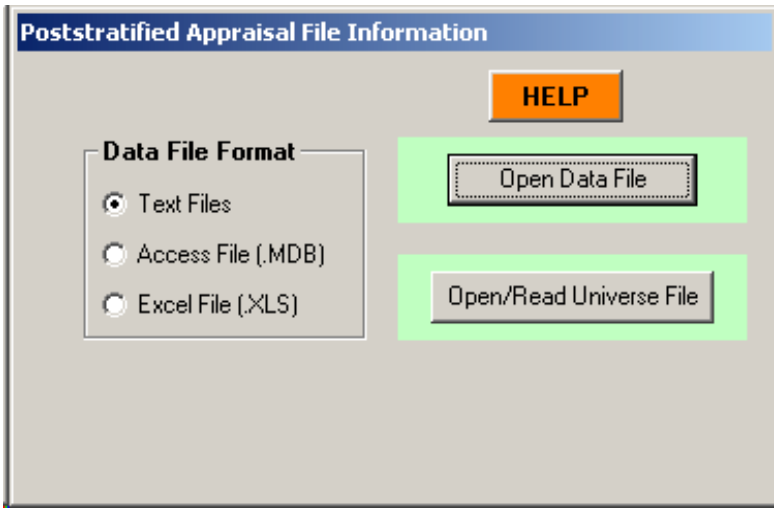
A portion of data set DATAPOST.TXT is shown next The data file contains 25 observations in each stratum.

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

Data file DATAPOST.TXT

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

After clicking on the **Open Data File** button and opening the data file (e.g., DATAPOST.TXT), the **Open/Read Universe File** button will appear on the following screen.

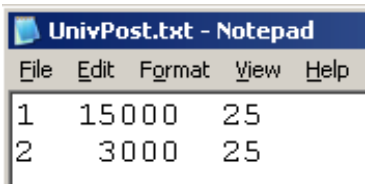


Universe/Sample Sizes File

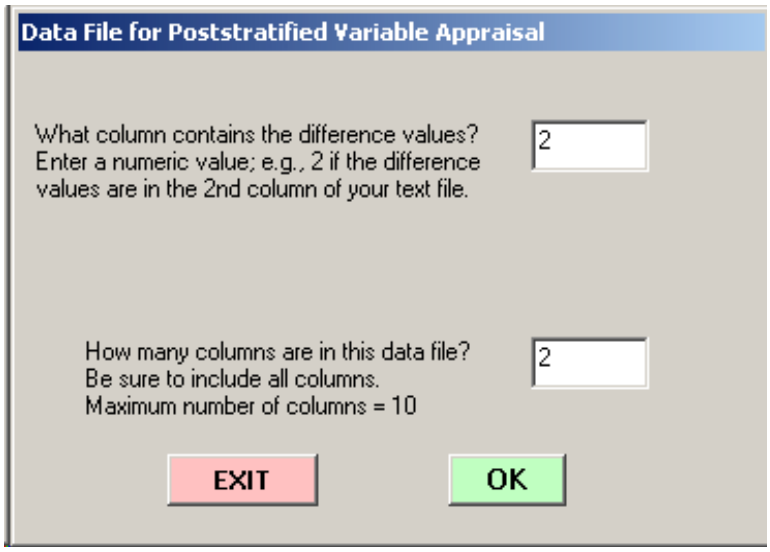
To open the file containing the universe/sample sizes, click on the **Open/Read Universe File** button and use the standard Windows "Open" file screen to locate the universe file (e.g., UnivPost.TXT, shown below). The format of this file is:

line counter, universe size, sample size

There is one line for each stratum in the sample. Values within a line can be separated by one or more spaces or by using the tab key. Commas are allowed in the universe and sample size values. After the universe file has been selected, the program will return to the input screen.



After opening the data files, the user is returned to the input screen. Click on **CONTINUE** to resume processing. The user will need to specify the column(s) containing the input data and the total number of columns in the data file using the screen shown next. After entering the column information, click on **OK**. The program will resume processing.



Input From an Access Database

The sample data and universe/sample size information must be stored in two tables within the same Access database. Select the name of the database containing the input tables 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 on the standard Windows "Open" file screen.

Universe/Sample Sizes Table

The name of the Access table containing the universe/sample size information for this illustration is UnivPost. The following table shows the table contents:

UnivPost : Table

Field1	Field2	Field3
1	15000	25
2	3000	25

Sample Data Table

The name of the Access table for this illustration is DATAPOST. 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.

DATAPOST : Table

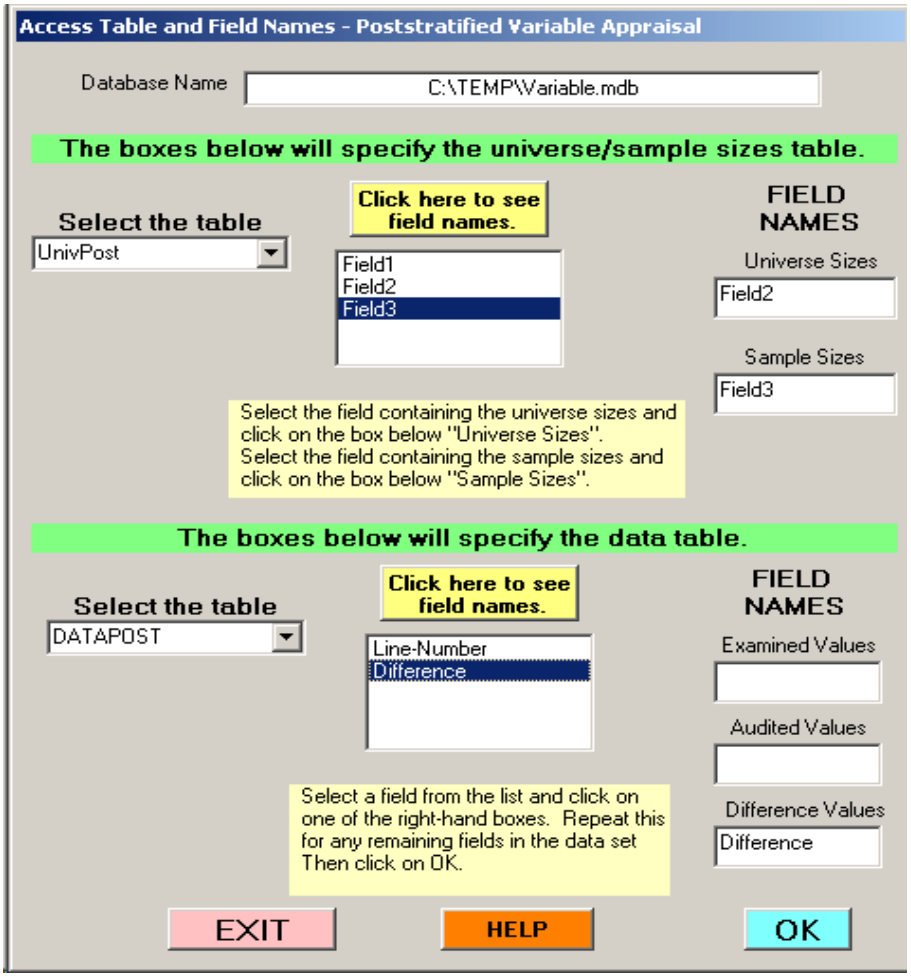
Line-Number	Difference
1	80
2	43
3	133
4	125
5	116

This data file contains 25 observations in each stratum. The last three rows of the first stratum and the first two rows of the second stratum are shown below:

23	102
24	69
25	76
26	354
27	328

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

After opening the database, the user will be asked to select the name of the table containing the universe/sample information and the table containing the sample data using the following form. For the universe/sample sizes table, click on the down arrow under **Select the table**. After selecting a table from the drop-down list (UnivPost for this illustration), click on **Click here to see field names**. This list will contain the field names for this table. To select the field names, click on the field name for the field containing the universe sizes ("Field2" in this illustration) and click on the box labeled "Universe Sizes." The field name will appear in this box. Repeat this procedure for the field containing the sample sizes ("Field3" in this illustration).



To select the sample data field names, first select the table (DATAPOST in this illustration) and click on the field name for the field containing the first piece of information in the input file ("Difference" in this illustration), then 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 return to the input screen. **NOTE**: When the user returns to the input screen, the data file format (Difference Values for this illustration) will be selected based on responses within the preceding **Access Table and Field Names** window. When using an Access input file, the user cannot change this data file format option after returning to the input screen.

Input From Excel Spreadsheets

With this option, the sample data and universe/sample sizes information must be stored in two Excel spreadsheets within the same Excel file. Select the name of the file containing both spreadsheets in the preceding **Open Data File** step. This file must have the standard Excel extension (.XLS). For this illustration, Excel file DATAPOST.XLS will be used.

Universe/Sample Size Spreadsheet

The contents of the spreadsheet containing the universe/sample sizes (named "Sizes" in this illustration) are shown below:

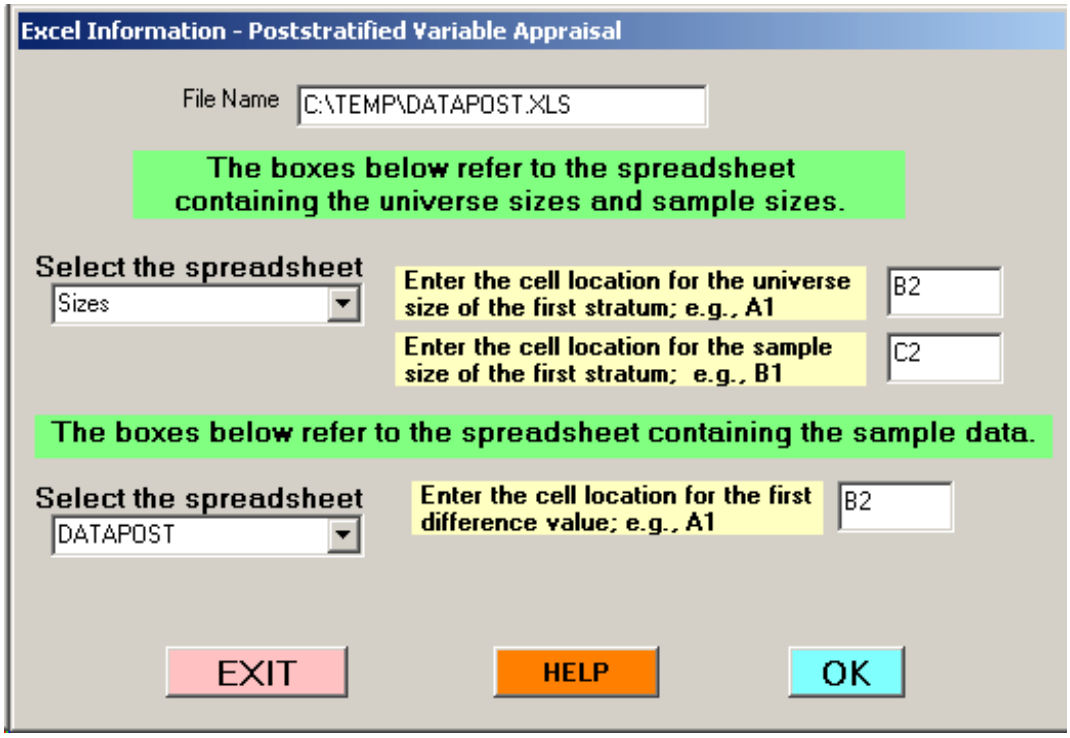
	A	B	C
1	Stratum	Universe Size	Sample Size
2	1	15000	25
3	2	3000	25

Sample Data Spreadsheet

The name of the second spreadsheet containing the sample data is DATAPOST in this illustration. The first five observations in this spreadsheet are shown below. The data file contains 25 observation in each stratum for a total of 51 rows (including the first row with labels). This particular file contains line numbers (1, 2, 3, . . .) in column A. The line numbers are optional.

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

After opening the data file, the user is returned to the input screen. Click on **CONTINUE** to resume processing. When using the Excel option, the following screen will be displayed. For this illustration, the various boxes should be filled in as shown. After entering the cell locations, click on **OK**. The program will resume processing.

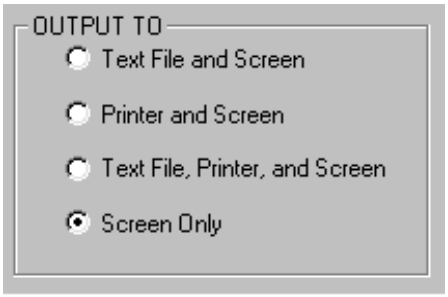


NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

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

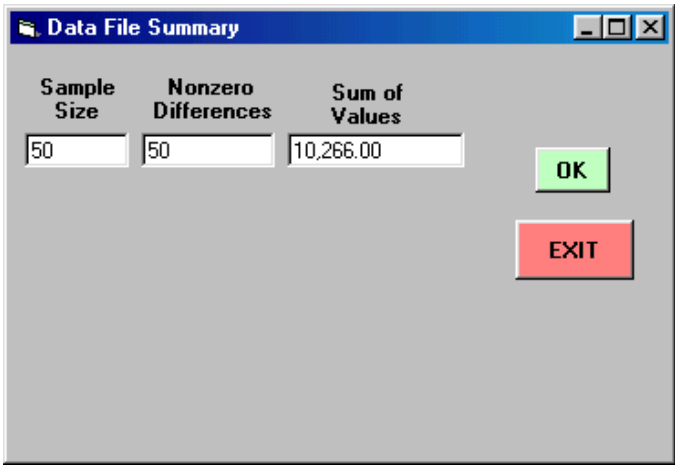


The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always concludes with a summary on the screen.

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.

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

For the examined, adjusted, and difference sections of the output, the following pieces of information will be displayed for each stratum. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

SAMPLE SIZE	The number of sample items belonging to this particular stratum (stratum sample size) or in the combined sample (overall sample size).
UNIVERSE SIZE	The number of universe items for this particular stratum (stratum universe size) or in the entire universe (overall universe size).
MEAN	The average value for the sample items appraised within a stratum. It is obtained by summing the sample items for this stratum and dividing the result by the number of sample items within this stratum.
STANDARD DEVIATION	A measurement of the variation of the sample items within a stratum about the stratum mean.
STANDARD ERROR (TOTAL)	A measurement of the variation of the estimated stratum universe total with respect to all possible estimated totals.
OVERALL STANDARD ERROR (TOTAL)	A measurement of the variation of the point estimate of the overall universe total with respect to all possible totals.
POINT ESTIMATE	For each stratum, the estimate of the stratum total. It is found by multiplying the sample mean and the number of universe items for this stratum. The overall point estimate is the sum of the strata point estimates.
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 stratum total and the actual stratum total. The precision amount is calculated by multiplying the stratum standard error by the appropriate factor ("Z" 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 The result of dividing the precision amount by the point estimate and stating the result as a percentage.

Z-VALUE USED The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

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

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004        POSTSTRATIFIED VARIABLE APPRAISAL         Time: 12:45
               AUDIT/REVIEW: Variable - Poststratification

                     DATA FILE USED: C:\TEMP\DATAPOST.TXT

           ----------------------D I F F E R E N C E----------------------
Stratum 1  SAMPLE SIZE / UNIVERSE SIZE                25             15,000
           MEAN                                    99.24
           STANDARD DEVIATION                      26.33
           STANDARD ERROR (TOTAL)              61,226.65
           POINT ESTIMATE                      1,488,600
```

```
                                  CONFIDENCE LIMITS
                                 80% CONFIDENCE LEVEL
           LOWER LIMIT                      1,410,135
           UPPER LIMIT                      1,567,065
           PRECISION AMOUNT                    78,465
           PRECISION PERCENT                     5.27%
           Z-VALUE USED                 1.281551565545

                                 90% CONFIDENCE LEVEL
           LOWER LIMIT                      1,387,891
           UPPER LIMIT                      1,589,309
           PRECISION AMOUNT                   100,709
           PRECISION PERCENT                     6.77%
           Z-VALUE USED                 1.644853626951

                                 95% CONFIDENCE LEVEL
           LOWER LIMIT                      1,368,598
           UPPER LIMIT                      1,608,602
           PRECISION AMOUNT                   120,002
           PRECISION PERCENT                     8.06%
           Z-VALUE USED                 1.959963984540

Stratum 2  SAMPLE SIZE / UNIVERSE SIZE            25            3,000
           MEAN                                 311.40
           STANDARD DEVIATION                    39.64
           STANDARD ERROR (TOTAL)            43,154.68
           POINT ESTIMATE                      934,200

                                  CONFIDENCE LIMITS
                                 80% CONFIDENCE LEVEL
           LOWER LIMIT                        878,895
           UPPER LIMIT                        989,505
           PRECISION AMOUNT                    55,305
           PRECISION PERCENT                     5.92%
           Z-VALUE USED                 1.281551565545

                                 90% CONFIDENCE LEVEL
           LOWER LIMIT                        863,217
           UPPER LIMIT                      1,005,183
           PRECISION AMOUNT                    70,983
           PRECISION PERCENT                     7.60%
           Z-VALUE USED                 1.644853626951

                                 95% CONFIDENCE LEVEL
           LOWER LIMIT                        849,618
           UPPER LIMIT                      1,018,782
           PRECISION AMOUNT                    84,582
           PRECISION PERCENT                     9.05%
           Z-VALUE USED                 1.959963984540

OVERALL    SAMPLE SIZE / UNIVERSE SIZE            50           18,000
           POINT ESTIMATE                    2,422,800
           STANDARD ERROR (TOTAL)               74,907
```

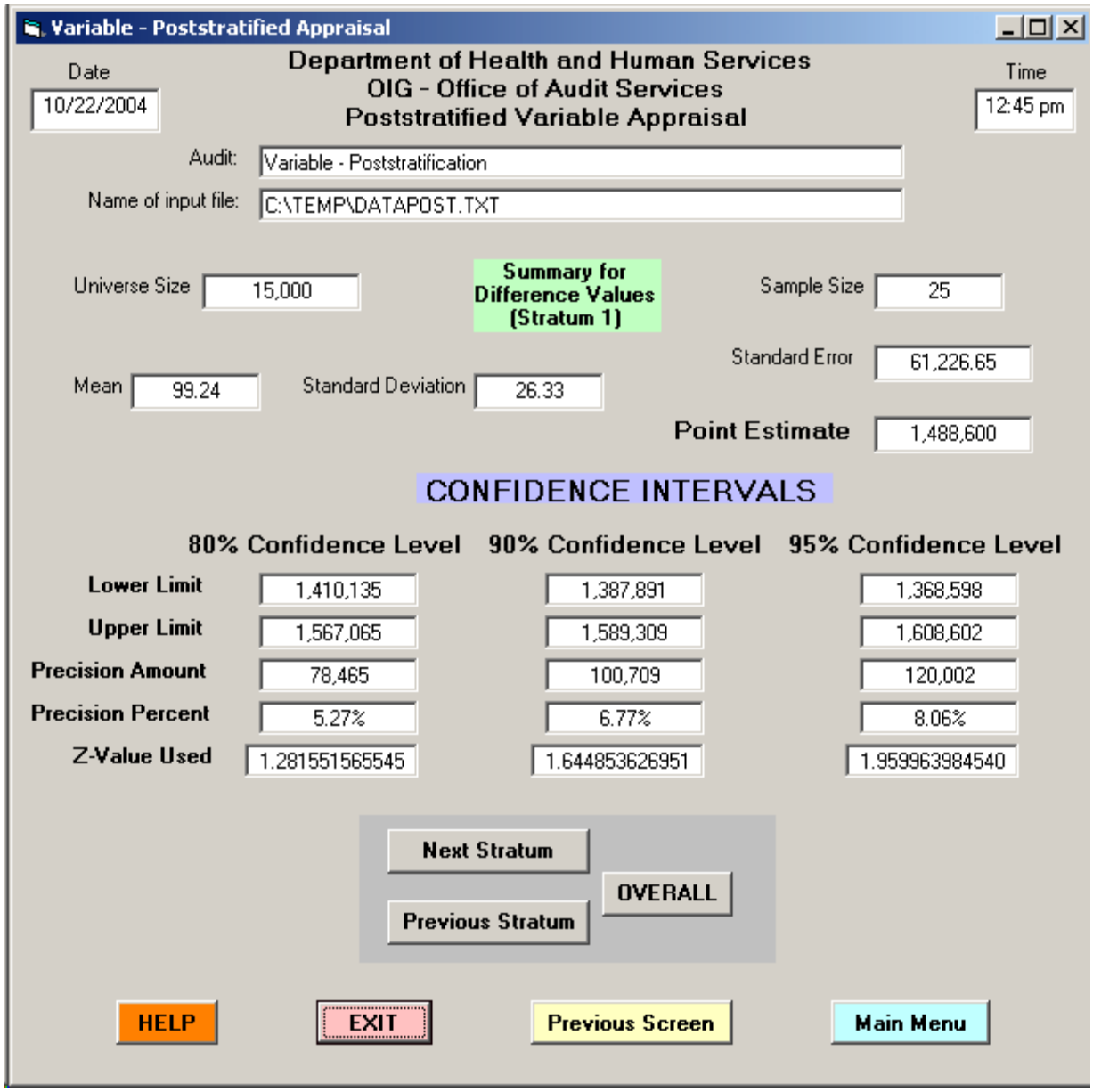
```
                                 CONFIDENCE LIMITS
                                80% CONFIDENCE LEVEL
LOWER LIMIT                               2,326,803
UPPER LIMIT                               2,518,797
PRECISION AMOUNT                             95,997
PRECISION PERCENT                              3.96%
Z-VALUE USED                        1.281551565545

                                90% CONFIDENCE LEVEL
LOWER LIMIT                               2,299,589
UPPER LIMIT                               2,546,011
PRECISION AMOUNT                            123,211
PRECISION PERCENT                              5.09%
Z-VALUE USED                        1.644853626951

                                95% CONFIDENCE LEVEL
LOWER LIMIT                               2,275,985
UPPER LIMIT                               2,569,615
PRECISION AMOUNT                            146,815
PRECISION PERCENT                              6.06%
Z-VALUE USED                        1.959963984540
```

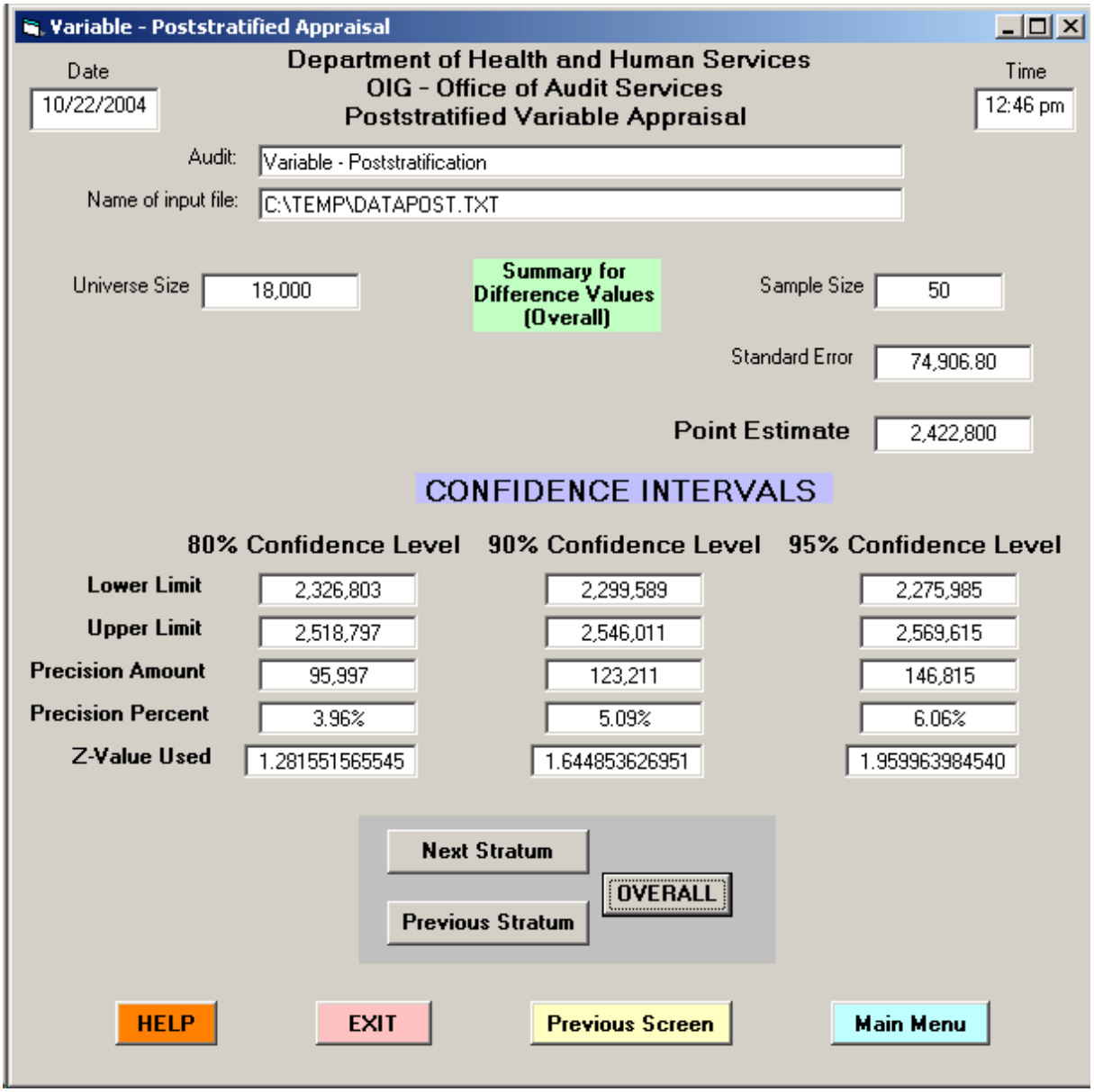
NOTE: Example is for illustrative purposes only. The sample sizes may not conform to the organization's minimum sample size standards.

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected is text file or printer. The screen below 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 form. For this illustration, the **Additional Summary Info** button is not visible since only the difference values were used 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**. The resulting summary screen immediately follows the summary for the first stratum.

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


Variable - Poststratified Appraisal
Date
10/22/2004
Department of Health and Human Services
OIG - Office of Audit Services
Poststratified Variable Appraisal
Time
12:45 pm
Audit:
Variable - Poststratification
Name of input file:
C:\TEMP\DATAPOST.TXT
Universe Size
15,000
Summary for Difference Values (Stratum 1)
Sample Size
25
Standard Error
61,226.65
Mean
99.24
Standard Deviation
26.33
Point Estimate
1,488,600
CONFIDENCE INTERVALS
80% Confidence Level
90% Confidence Level
95% Confidence Level
Lower Limit
1,410,135
1,387,891
1,368,598
Upper Limit
1,567,065
1,589,309
1,608,602
Precision Amount
78,465
100,709
120,002
Precision Percent
5.27%
6.77%
8.06%
Z-Value Used
1.281551565545
1.644853626951
1.959963984540
Next Stratum
Previous Stratum
OVERALL
HELP
EXIT
Previous Screen
Main Menu

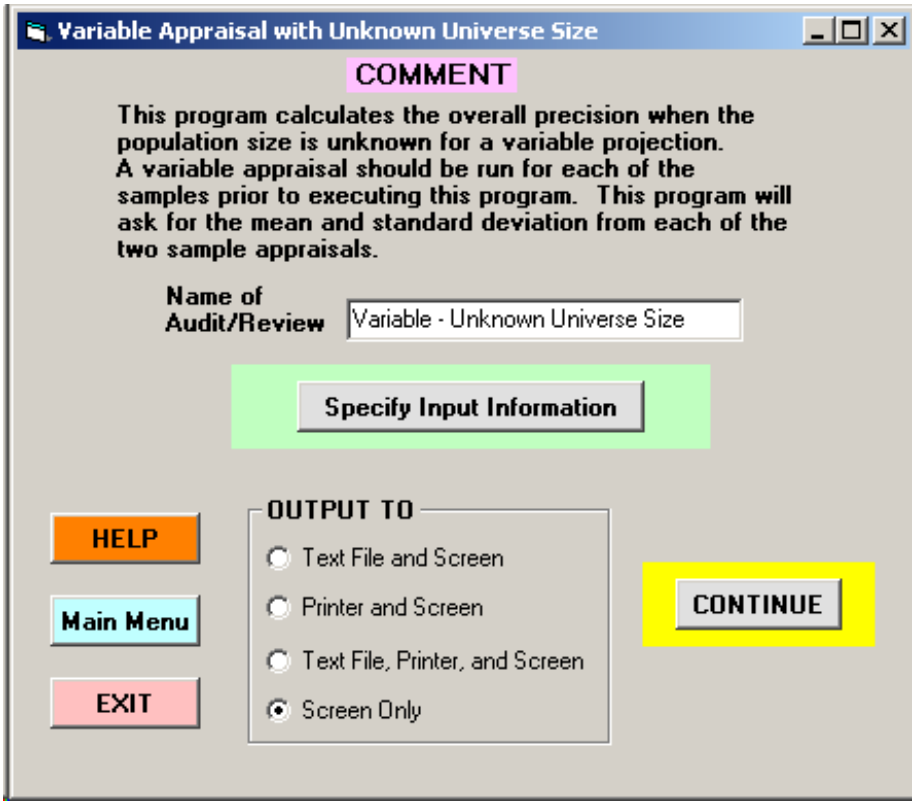

Variable - Poststratified Appraisal
Date
10/22/2004
Department of Health and Human Services
OIG - Office of Audit Services
Poststratified Variable Appraisal
Time
12:46 pm
Audit:
Variable - Poststratification
Name of input file:
C:\TEMP\DATAPOST.TXT
Universe Size
18,000
Summary for Difference Values (Overall)
Sample Size
50
Standard Error
74,906.80
Point Estimate
2,422,800
CONFIDENCE INTERVALS
80% Confidence Level
90% Confidence Level
95% Confidence Level
Lower Limit
2,326,803
2,299,589
2,275,985
Upper Limit
2,518,797
2,546,011
2,569,615
Precision Amount
95,997
123,211
146,815
Precision Percent
3.96%
5.09%
6.06%
Z-Value Used
1.281551565545
1.644853626951
1.959963984540
Next Stratum
OVERALL
Previous Stratum
HELP
EXIT
Previous Screen
Main Menu

UNKNOWN UNIVERSE SIZE

Purpose

This program calculates the overall precision for a population whose size is unknown. This program requires that two samples have already been taken and appraised. One sample was used to estimate the population size and the other sample was taken to estimate one or more variable characteristics. The two samples must be appraised prior to executing this module, since this program will ask for the mean and standard deviation of each sample.

Input Screen

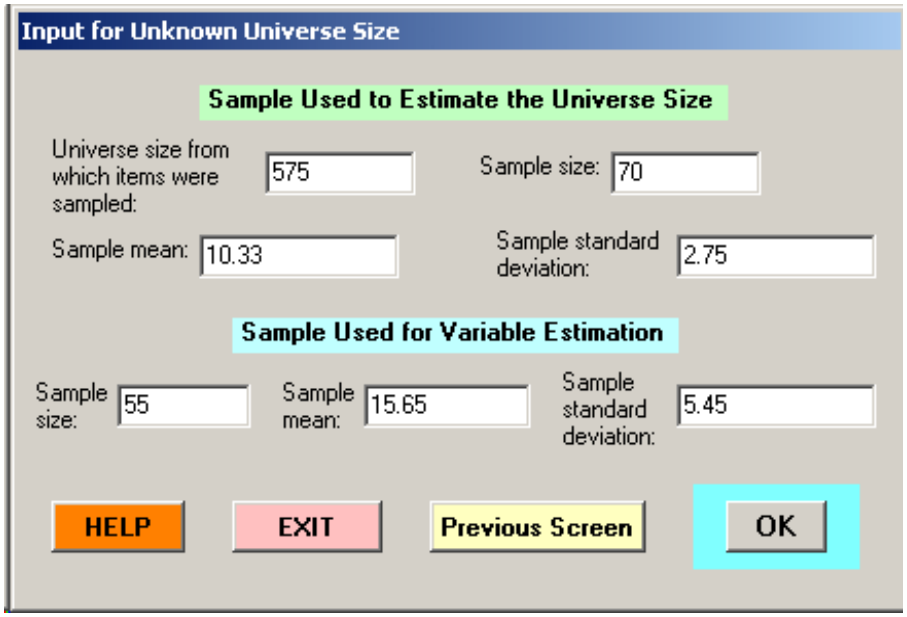


Name of audit/review

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.

Specify Input Information

The initial screen contains a warning to the user that prior appraisals of the two samples must be taken before running this module. To enter input information, click on **Specify Input Information**. The following form will appear:



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

Sample Used to Estimate the Universe Size

Universe size from which items were sampled

The universe of interest is a subset of some other universe. This larger sampling frame could be file drawers or pages in a check register. What the user has done is sampled from this universe (e.g., file drawers) and counted the number of items that meet the criteria for sample selection in the other sample. The results for each unit in this sample (e.g., file drawers) is entered into a data file for use in the variable appraisal program. The total number of sample units (e.g., file drawers) in this universe must be known.

Sample size

The user enters the number of sampling units (e.g., file drawers) drawn from the universe entered above.

Sample mean

The variable appraisal program will generate a mean for the sample. The user should enter that mean value.

Sample standard deviation

The variable appraisal program will generate a standard deviation for the sample. The user should enter that standard deviation value.

Sample Used for Variable Estimation

Sample size

The user has drawn a second sample of items that meet the criteria for review. The user should enter the size of this second sample.

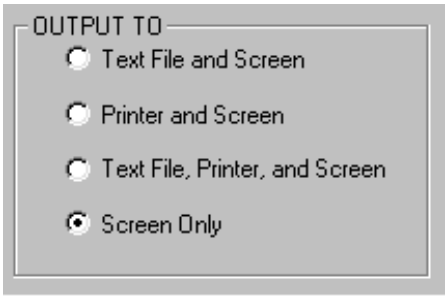
Sample mean

The variable appraisal program previously used with this sample generated a sample mean. The user should enter that mean value.

Sample standard deviation

The variable appraisal program previously used with this sample generated a sample standard deviation. The user should enter that standard deviation value.

Output Options



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or 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.

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.

Program Output

Based on the data given by the user, the sections of information will be generated by the program. The first section displays the information provided by the user.

UNIVERSE	The size of the universe from which the sample was drawn to estimate the population size.
SAMPLE	The two sample sizes that were drawn for this evaluation.
MEAN	The two mean values entered by the user.
STANDARD DEVIATION	The two standard deviation values entered by the user.

The second section of the output displays the results of the estimation. The precision information is given at the two-sided 80%, 90%, and 95% confidence levels.

POINT ESTIMATE	The estimate of the universe total.
STANDARD ERROR	A measurement of the variation of the point estimate of the total with respect to all possible totals for this universe and sample sizes.
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. 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	The result of dividing the precision amount by the point estimate.
Z-VALUE USED	The standard normal percentile value used to construct the confidence interval.

Output to a Text File or Printer

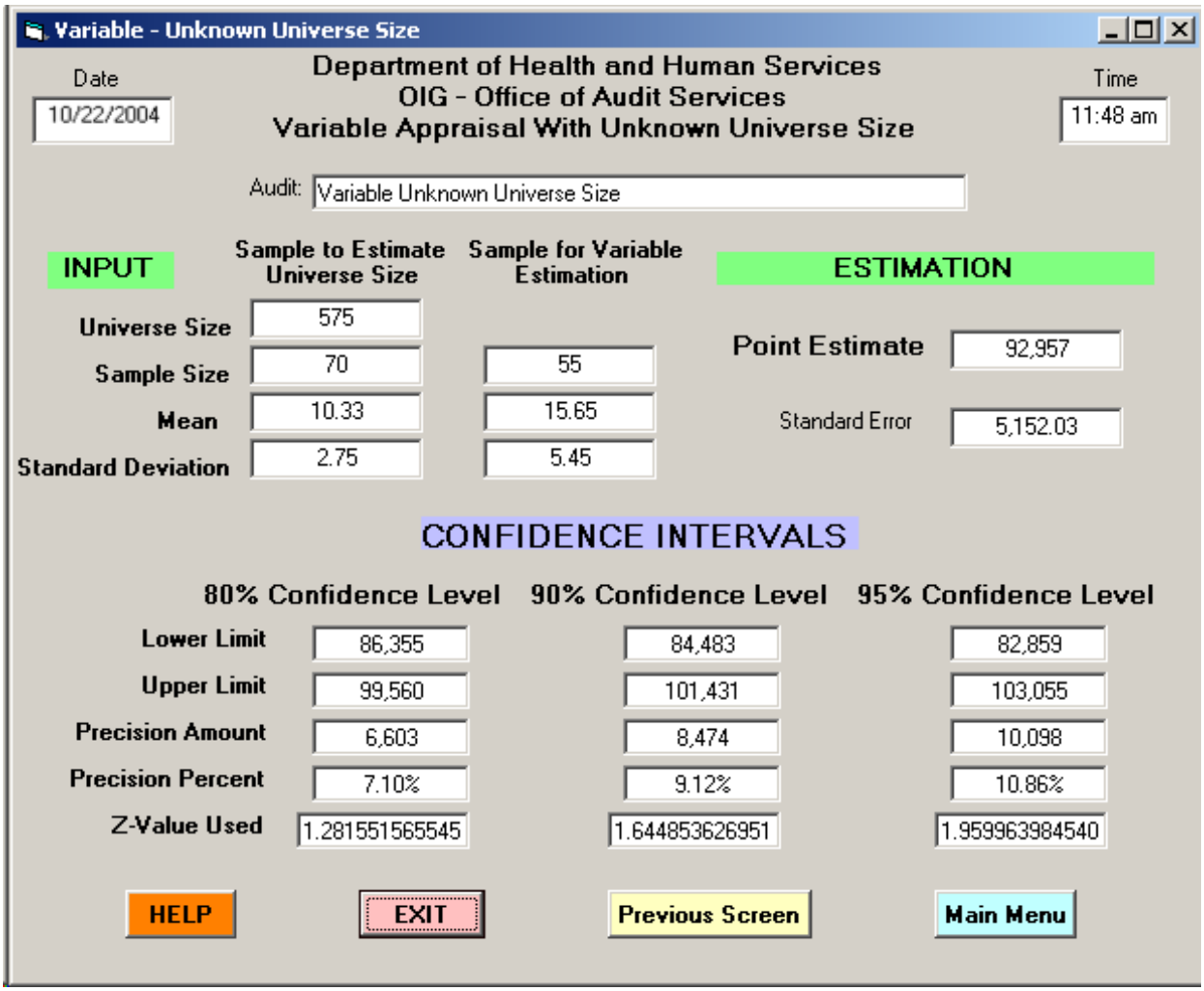
Using the specified sample results in the initial screen, the appraisal results were stored in C:\TEMP\OUTUNKNOWN.TXT, shown next. The printer output is identical.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004   VARIABLE APPRAISAL WITH UNKNOWN UNIVERSE SIZE   Time: 11:48
                   AUDIT/REVIEW: Variable Unknown Universe Size

   = = = = = = = = = = = = = = = = I N P U T = = = = = = = = = = = = = = =
                                     SAMPLE TO            SAMPLE FOR
                                  ESTIMATE POPULATION    VARIABLE ATTRIBUTE
   UNIVERSE                                575
   SAMPLE                                   70                    55
   MEAN                                     10.33                 15.65
   STANDARD DEVIATION                        2.75                  5.45

   = = = = = = = = = = = = =  E S T I M A T I O N  = = = = = = = = = = = = =
                     80% CONFIDENCE      90% CONFIDENCE      95% CONFIDENCE
   POINT ESTIMATE            92,957              92,957              92,957
   STANDARD ERROR             5,152               5,152               5,152
   LOWER LIMIT               86,355              84,483              82,859
   UPPER LIMIT               99,560             101,431             103,055
   PRECISION AMOUNT           6,603               8,474              10,098
   PRECISION PERCENT          7.10%               9.12%              10.86%
   Z-VALUE USED       1.281551565545      1.644853626951      1.959963984540
```

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

Output to Screen

The program always concludes with a summary on the screen, even if the output option selected is text file or printer. The screen below is the output screen for this illustration. The user supplied input is on the left side (labeled **INPUT**) and the remainder of the output form contains the appraisal results.



Variable - Unknown Universe Size

Date: 10/22/2004

Department of Health and Human Services
OIG - Office of Audit Services
Variable Appraisal With Unknown Universe Size

Time: 11:48 am

Audit: Variable Unknown Universe Size

INPUT	Sample to Estimate Universe Size	Sample for Variable Estimation
Universe Size	575	
Sample Size	70	55
Mean	10.33	15.65
Standard Deviation	2.75	5.45

ESTIMATION	
Point Estimate	92,957
Standard Error	5,152.03

CONFIDENCE INTERVALS

	80% Confidence Level	90% Confidence Level	95% Confidence Level
Lower Limit	86,355	84,483	82,859
Upper Limit	99,560	101,431	103,055
Precision Amount	6,603	8,474	10,098
Precision Percent	7.10%	9.12%	10.86%
Z-Value Used	1.281551565545	1.644853626951	1.959963984540

HELP EXIT Previous Screen Main Menu

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

Sample Size Determination

OVERVIEW

VARIABLE SAMPLE SIZE DETERMINATION
- - **Unrestricted - Using Reported Amounts**
- - **Unrestricted - Using Estimated Error Rate**
- - **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 included 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 (Using Reported Amounts) 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 Unrestricted (Using Estimated Error Rate) module determines an optimum sample size by first estimating the mean and standard deviation of the difference amounts using the reported amounts and an estimated error rate. 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 when obtaining 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.

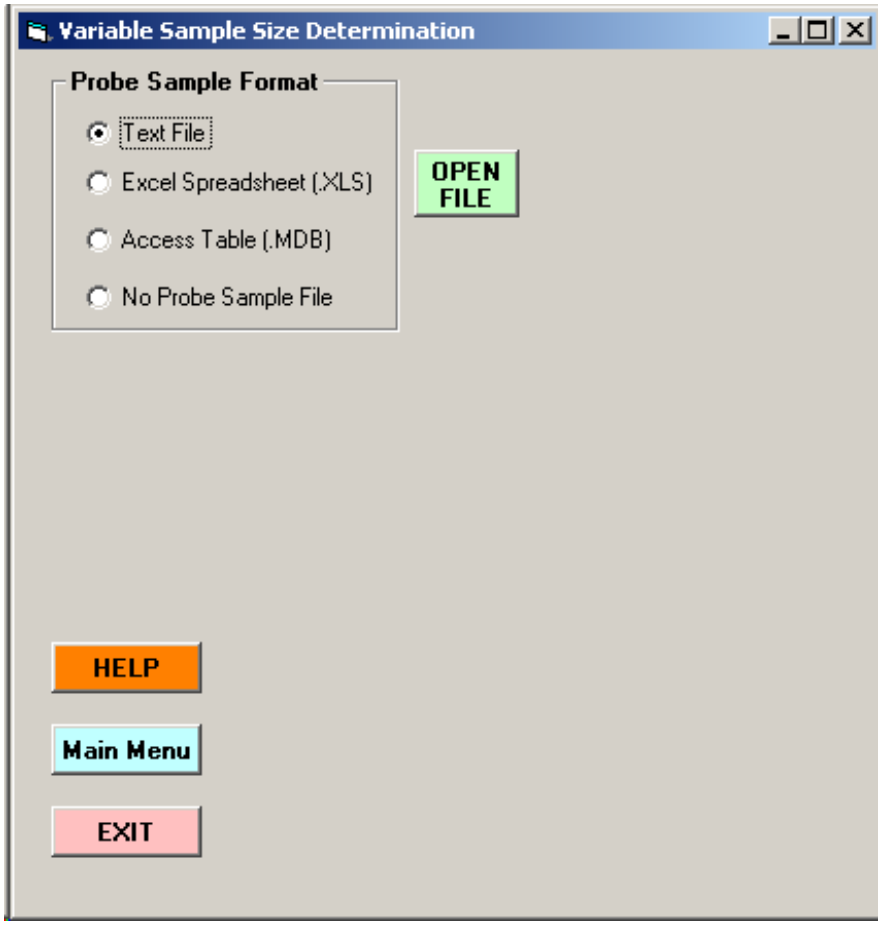
VARIABLE SAMPLE SIZE DETERMINATION UNRESTRICTED - Using a Probe Sample

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:

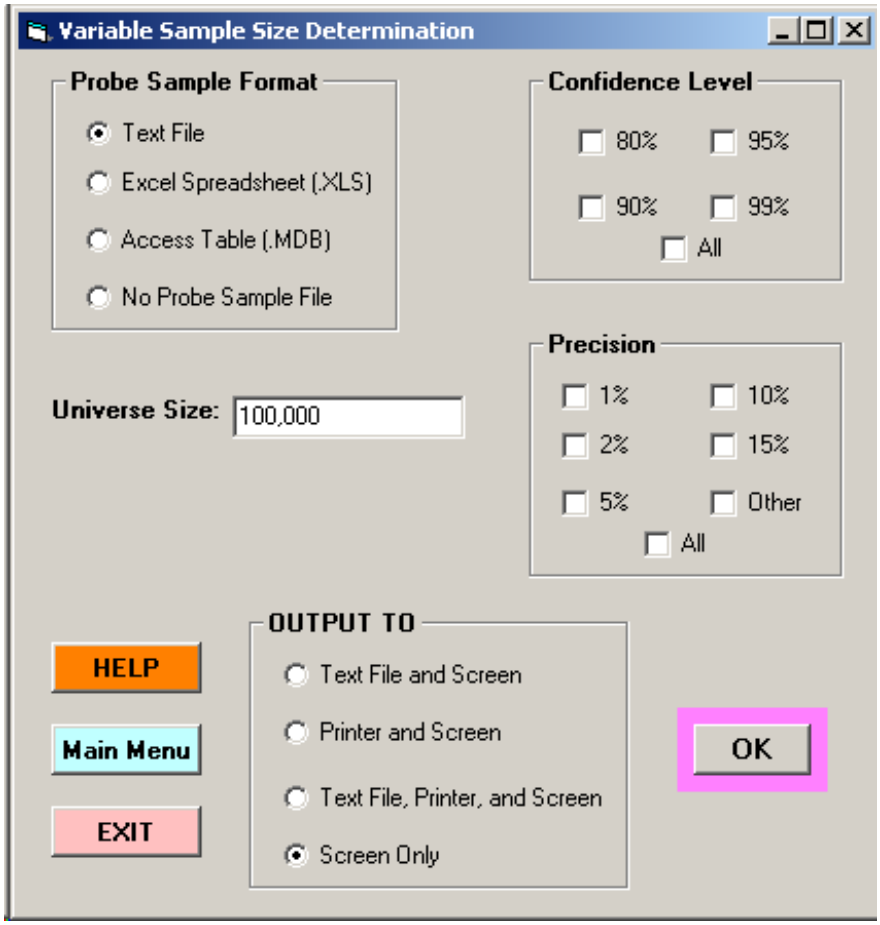


Probe Sample

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.

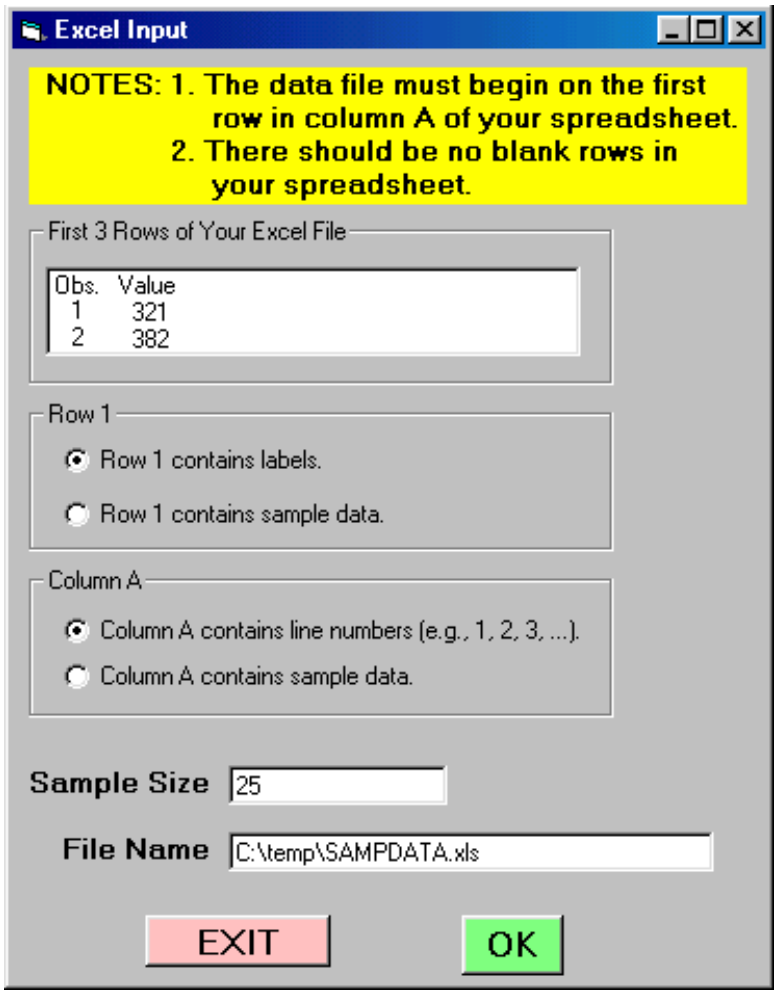
321
382
453
459
343
388
313
420
407
395
441
448
447
333
357
395
477
391
356
368
376
350
461
472
447

The probe text file in C:\TEMP\SAMPDATA.TXT.

(mean = 400, standard deviation = 50)

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.


Excel Input
NOTES: 1. The data file must begin on the first row in column A of your spreadsheet.
2. There should be no blank rows in your spreadsheet.
First 3 Rows of Your Excel File
Obs. Value
1 321
2 382
Row 1
Row 1 contains labels.
Row 1 contains sample data.
Column A
Column A contains line numbers (e.g., 1, 2, 3, ...).
Column A contains sample data.
Sample Size
25
File Name
C:\temp\SAMPDATA.xls
EXIT
OK

	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

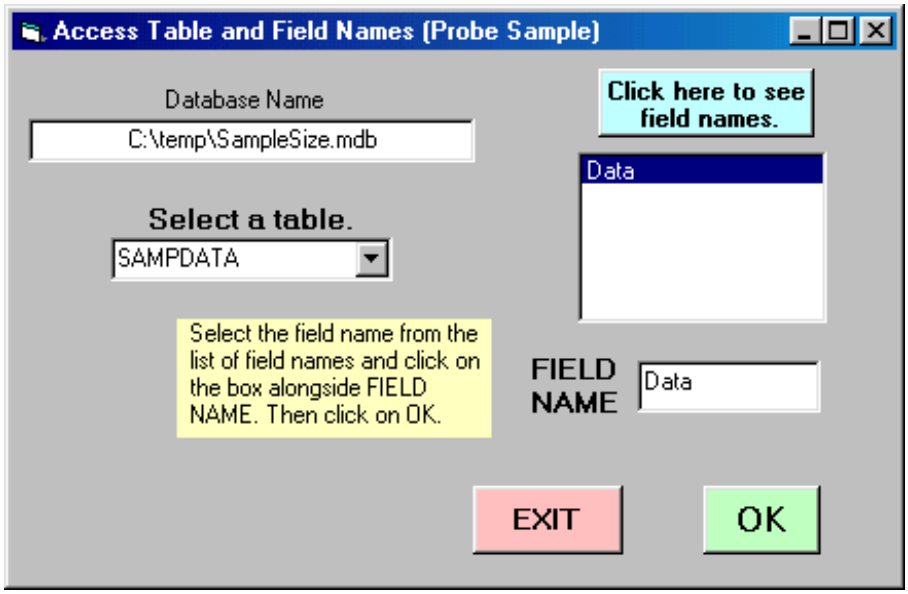
The probe Excel file in C:\TEMP\SAMPDATA.XLS.

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 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), click on **Click here to see field names**. This list will contain the field names for this table. 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 File Used

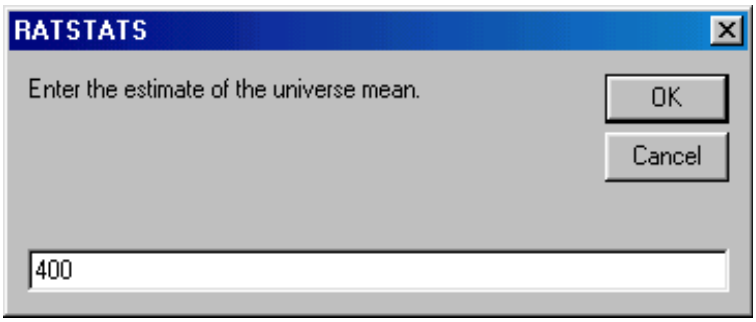
The user has the option of not using a probe sample in this analysis. Using this option, the user will be asked to specify the anticipated mean for the sample. If the mean is not known, the best estimate of the mean may be used. Other sources of data, such as prior reviews, may provide assistance in 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.

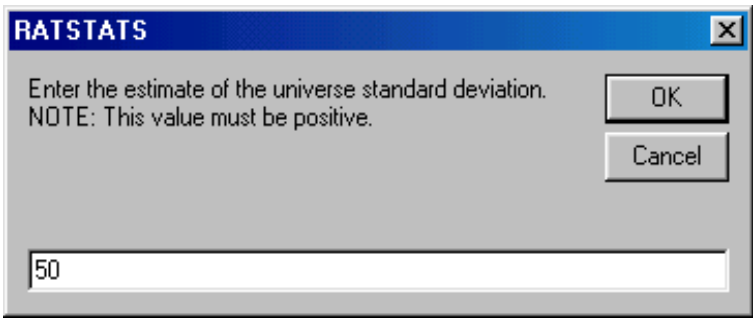
When this option is selected the user will see the following message:

By clicking on **OK**, the user will return to the full input screen for this program, shown in the **Probe Sample in a Text File** section.

After selecting the desired confidence levels, precision percentages, and output types, the user will click on **OK** to continue. 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

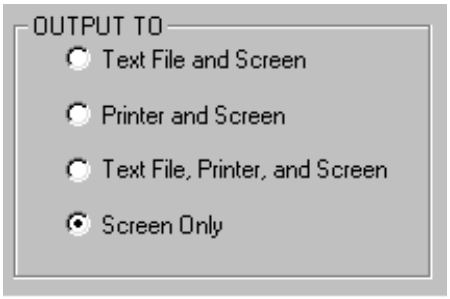
The user may select any combination of the following confidence levels: 80%, 90%, 95%, and 99%. Selecting all four confidence levels can be done by clicking on the "All" option. If none of the options are selected, the program will use the default "All" option.

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 output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always finishes with the screen output.

Explanation of Output

The output for each cell in the output table will consist either the necessary sample size or 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

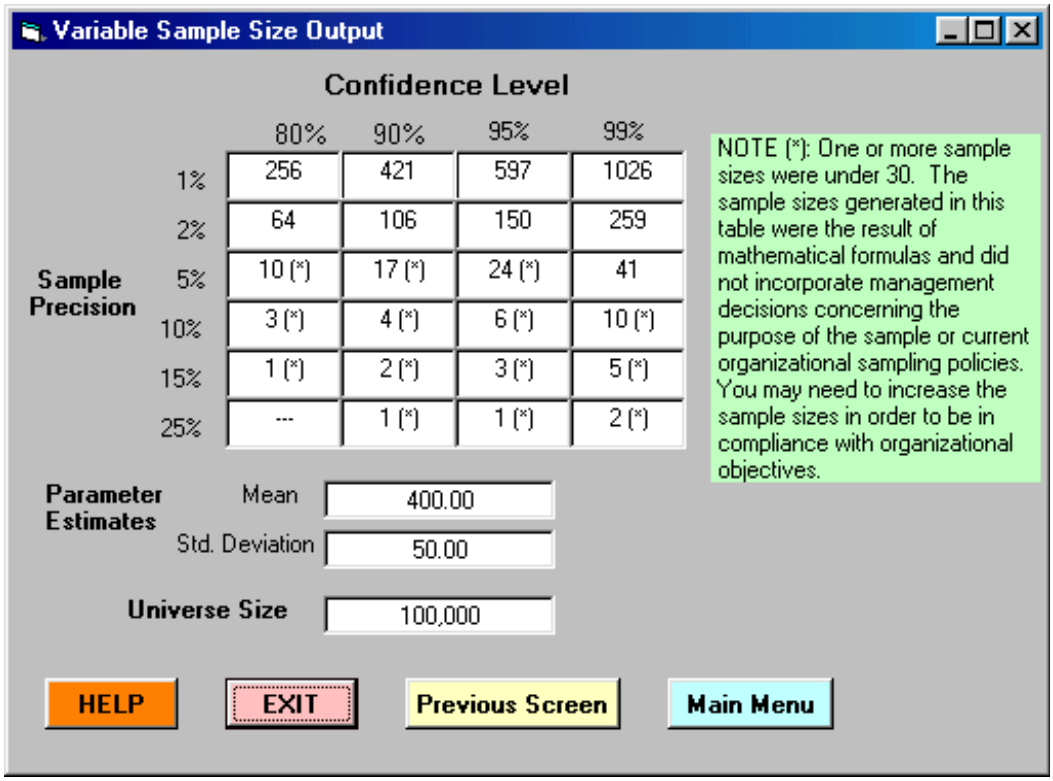
Using the probe sample in C:\TEMP\SAMPDATA.TXT, the sample sizes were saved in C:\TEMP\OUTSIZES.TXT, shown below. The printer output is identical. If any of the sample sizes are under 30, the note shown following the calculated sample sizes is the final portion of the program output.

```
                DEPARTMENT OF HEALTH & HUMAN SERVICES
                    OIG - OFFICE OF AUDIT SERVICES
Date: 10/22/2004      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


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.
```

NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

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.



NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

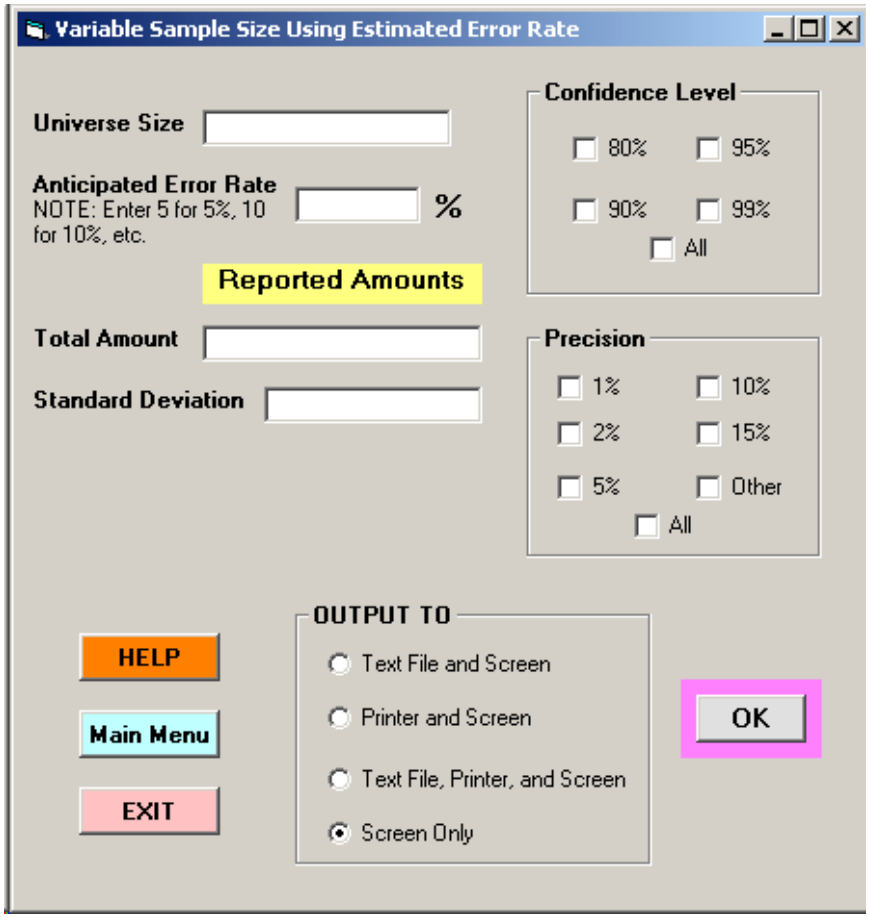
VARIABLE SAMPLE SIZE DETERMINATION UNRESTRICTED - Using Expected Error Rate

Purpose

This program allows the user to estimate sample sizes for specified precision percentages and specified confidence levels. The program determines an optimum sample size by first estimating the mean and standard deviation of the difference amounts using the reported amounts and an estimated error rate.

Input Screen

The input screen for this program is shown below:



Assumptions

This procedure estimates the mean and standard deviation of the difference (error) amounts by assuming (1) any item found to be in error is 100% in error and (2) the mean and standard deviation of the *nonzero* error amounts is the same as the mean and standard deviation of the reported (examined) amounts. Even though these assumptions may not be entirely true, this procedure will often give more reliable sample size estimates than those obtained using the Variable Unrestricted (Using Reported Amounts) module since the expected number of zero values in the error population is factored into the sample size calculation. The mean and standard deviation of the error amounts are estimated by assuming the percentage of nonzero errors in the error population is equal to the expected error rate (one of the input values) and the nonzero errors resemble the reported amounts; that is, the mean and standard deviation of the nonzero errors are equal to the mean and standard deviation of the reported amounts.

Screen Input

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.

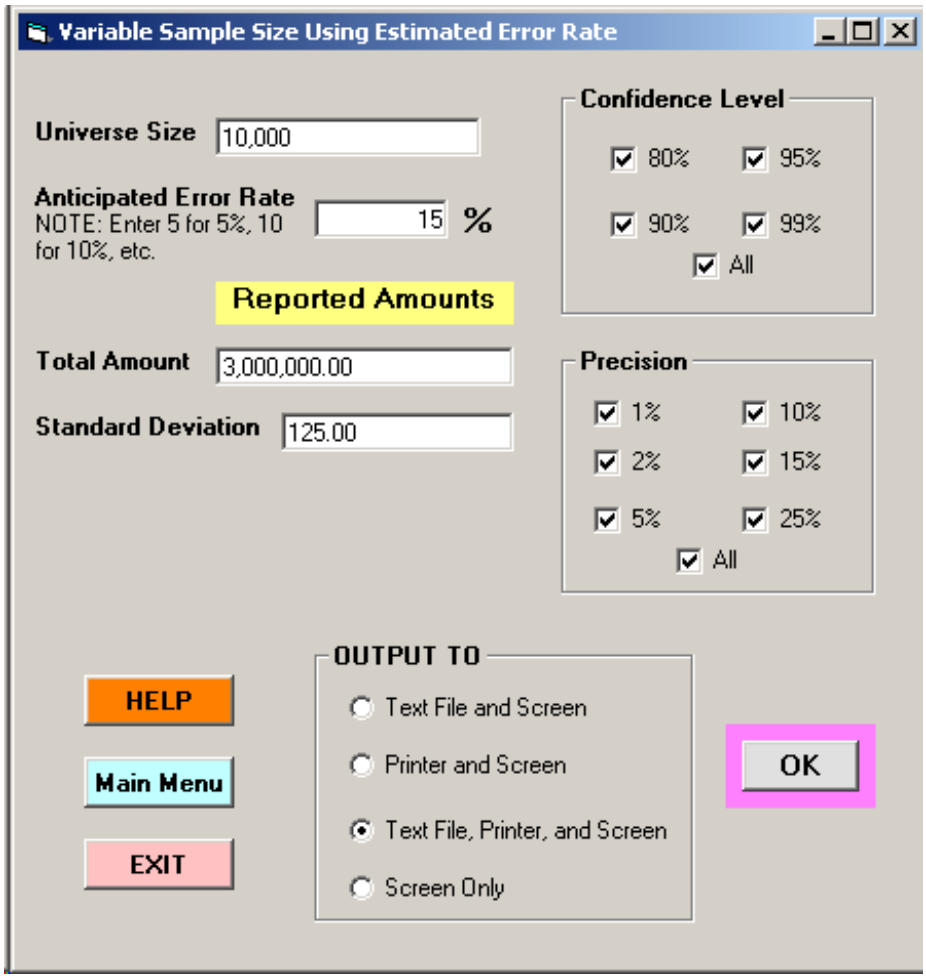
Precision

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.

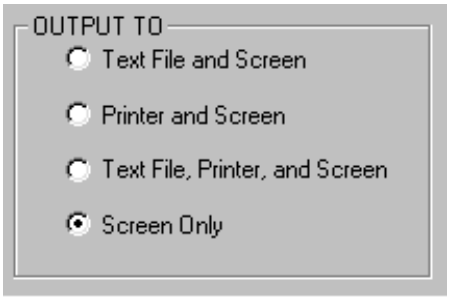
A confidence interval for the universe mean is obtained by adding and subtracting an amount (say, E) to/from the sample mean. The value of E is determined using the universe size, sample size, standard deviation, and the selected confidence level. The sample precision percentage is calculated as E divided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

Example

The estimated error rate is 15% for a universe of 10,000 transactions. The total reported amount is $3,000,000 and the standard deviation of the reported amounts is $125. Consequently, the mean reported amount is $300. Of interest is the required sample size necessary in order to obtain plus or minus 15% using a 90% confidence level. The corresponding input screen follows where 25% was specified for the "Other" precision level.



Program Output



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or screen only. The program always finishes with the screen output.

Explanation of Output

The output for each cell in the output table will consist either the necessary sample size or the text "- - -." The necessary sample size is the number of sample items necessary to obtain the specified sample precision at the specified confidence level. In this illustration a sample size of 287 is necessary to obtain a point estimate having a precision percentage of plus or minus 15% 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 estimated mean and standard deviation of the difference (error) values. For this illustration, the estimated mean and standard deviation are $45.00 and $117.55, respectively.

Output to a Text File or Printer

The output results were saved in C:\TEMP\OUTSAMPLE.TXT, shown on the next page. The printer output is identical. Notice that obtaining precision levels of 10% or less will require extremely large sample sizes.

```
                    DEPARTMENT OF HEALTH & HUMAN SERVICES
                        OIG - OFFICE OF AUDIT SERVICES
Date: 12/22/2004            Sample Size Determination            Time: 10:14


                             Confidence Level
                 80%         90%           95%           99%
            1%   9181        9486          9633          9784
            2%   7370        8219          8676          9188
Precision   5%   3095        4248          5119          6443
Level      10%   1008        1559          2077          3117
           15%   474         758           1044          1675
           25%   176         287           403           675


Universe Size: 10,000

Anticipated Error Rate: 15%

Reported Amounts - - Total Amount: 3,000,000.00

                     Standard Deviation: 125.00


Difference Values - - Estimated Mean: 45.00

                      Estimated Standard Deviation: 117.55
```

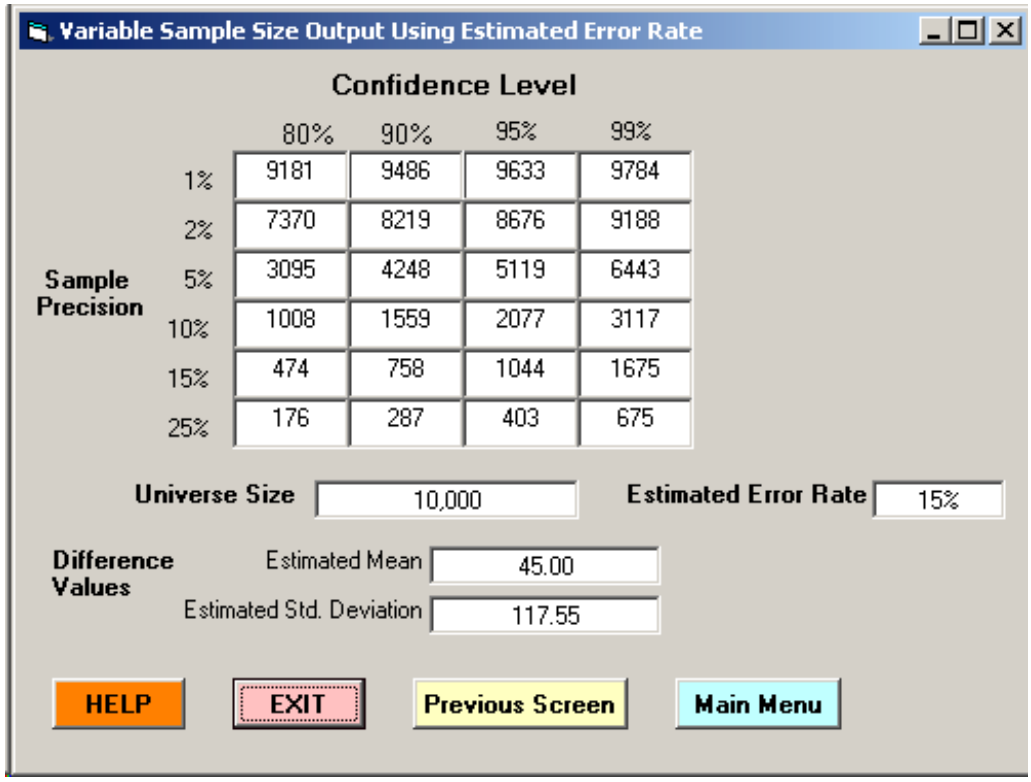
NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

If any of the sample sizes are under 30, the note shown below will be the final portion of the program 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. A note will appear on the right side of this screen whenever one or more of the sample sizes are under 30.



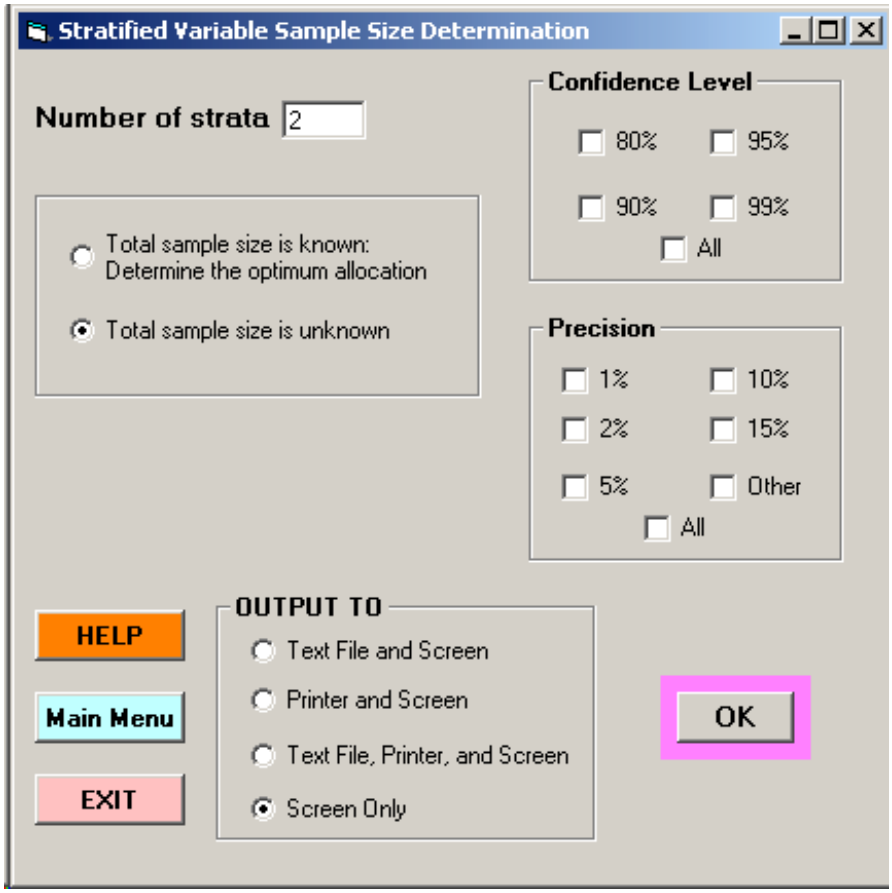
VARIABLE SAMPLE SIZE DETERMINATION STRATIFIED

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 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

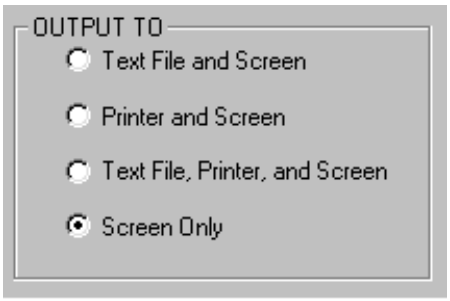
Confidence Level

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

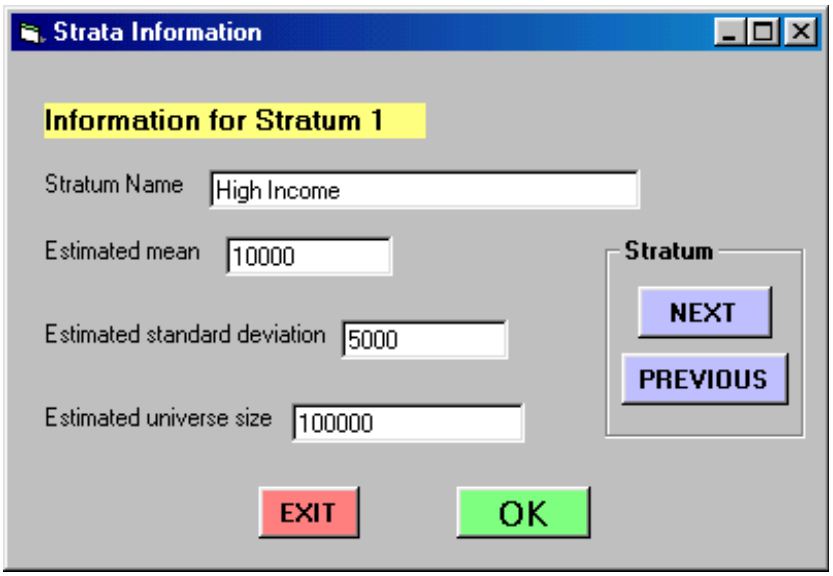
Sample precision is measured as a percentage. The available precision percentages are 1%, 2%, 5%, 10%, 20%, and "Other." By selecting "All," the user will obtain the sample sizes for the first five precision percentages. When selecting "Other," the user will be prompted to enter the desired precision 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 a text file and screen; a printer and screen; a text file, printer, and screen; or 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 either the necessary sample size or 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 111 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 estimated mean, estimated standard deviation, and 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: 10/22/2004           Sample Size Determination                Time: 10:40


              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%
            1%   8264         13492          18975          32030
            2%   2089         3432           4860           8345
Precision   5%   335          553            784            1352
Level      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.
```

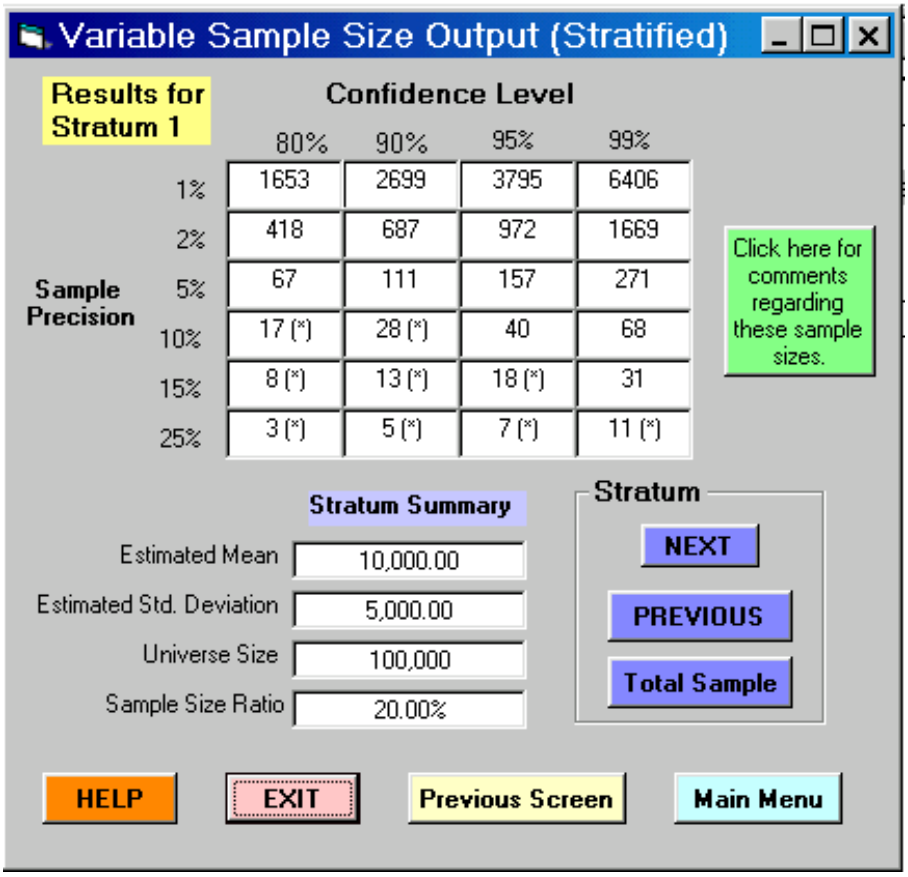
If any of the calculated samples sizes exceeds the corresponding universe size, the program will conclude with the following reminder:

```
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.
```

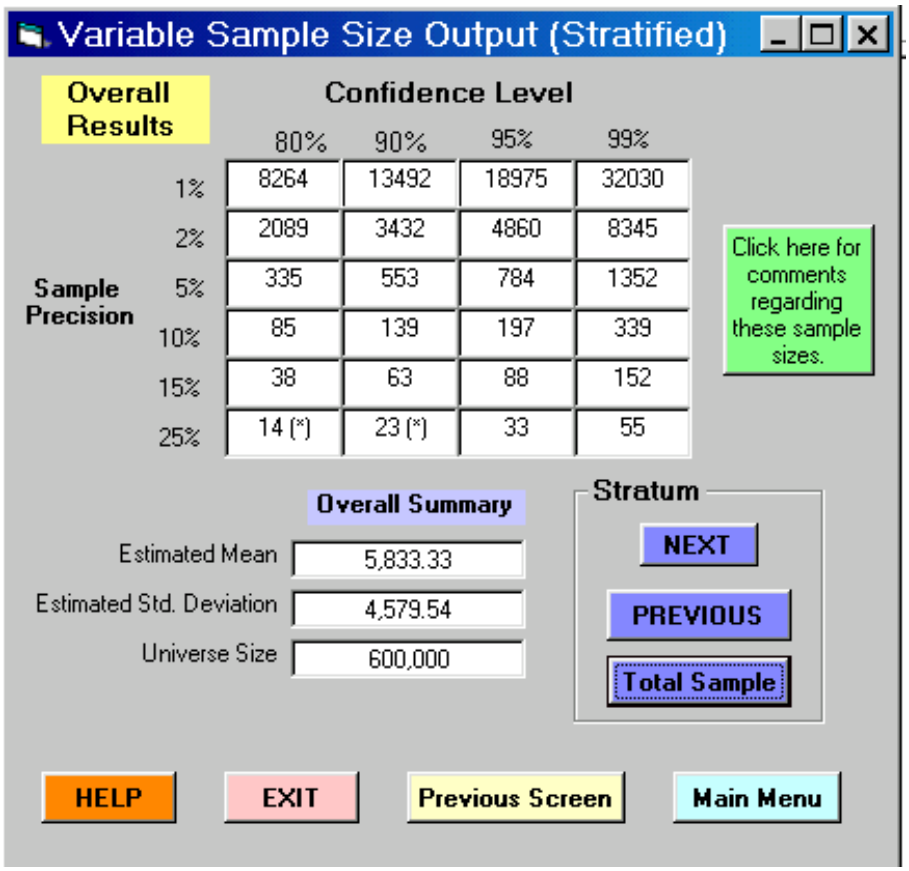
NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

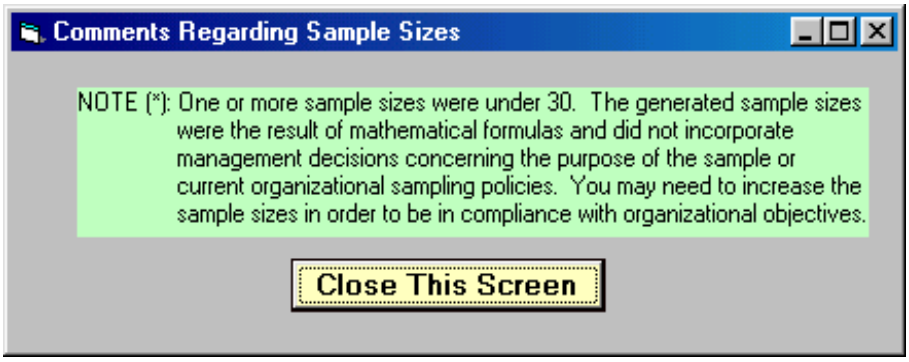
Output to Screen

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 **Output to a Text File or Printer** section.



NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.





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: 10/22/2004            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%
Sample Precision  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%
```

> **NOTE**: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

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:

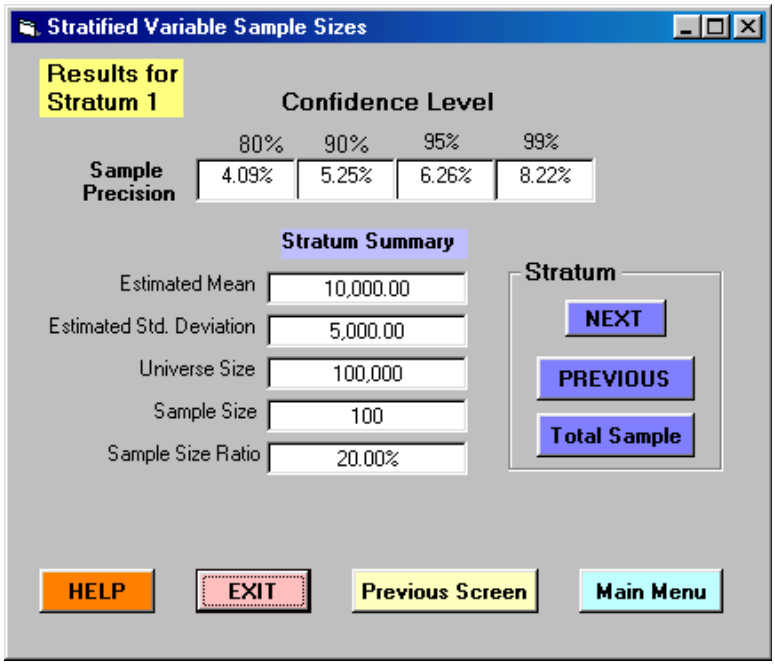
```
NOTE (!):   Due to rounding, the total sample size calculated
            (499) does not equal the sample size requested (500).
```

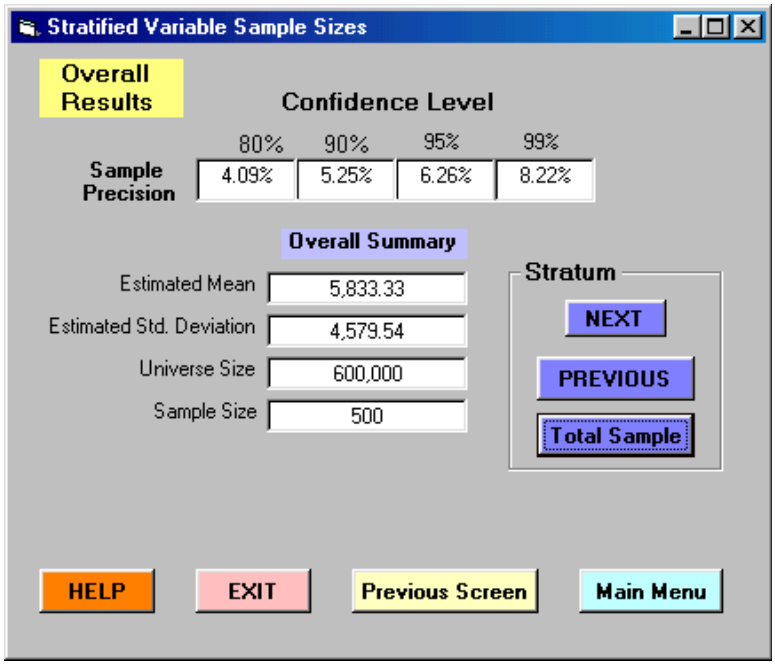
Output to Screen

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.





NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

ATTRIBUTE SAMPLE SIZE DETERMINATION

Purpose

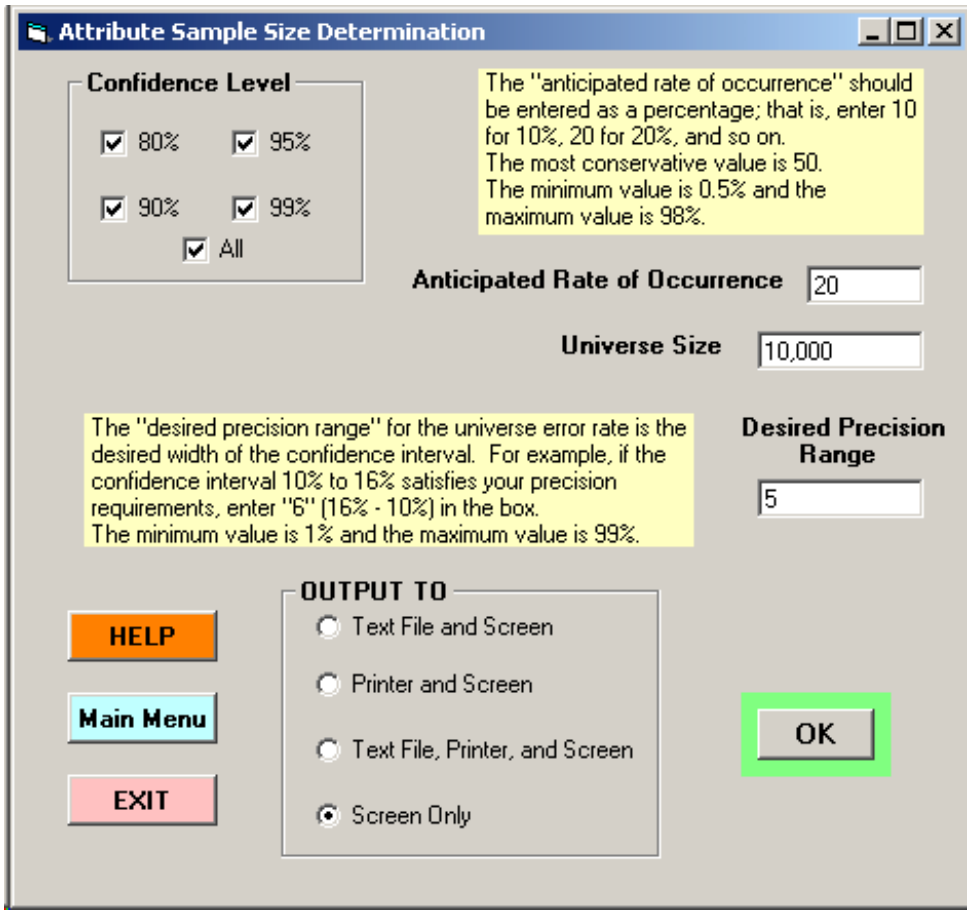
This program determines the sample size for an attribute simple random sample. The sample size is determined to provide for a specified degree of precision (using the desired width of the confidence interval) at four levels of confidence (80%, 90%, 95%, and 99%). The resulting sample sizes are the smallest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

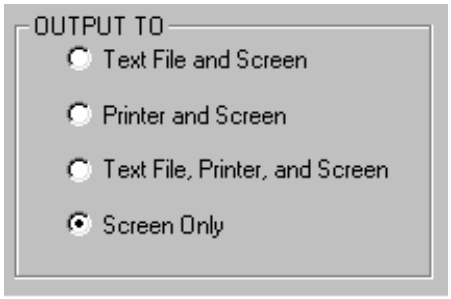
Universe Size

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 15% satisfies the user's precision requirements, the response to this query would be "5."

Program Output



The output options are a text file and screen; a printer and screen; a text file, printer, and screen; or 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 either the necessary sample size or 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 680 is necessary to obtain a confidence interval having a width of 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 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: 10/22/2004   Sample Size Determination          Time: 8:46


                             Confidence Level
                 80%         90%           95%           99%
Sample Size      439         680           926         1,480


Anticipated Rate of Occurrence: 20%

Desired Precision Range: 5%

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:

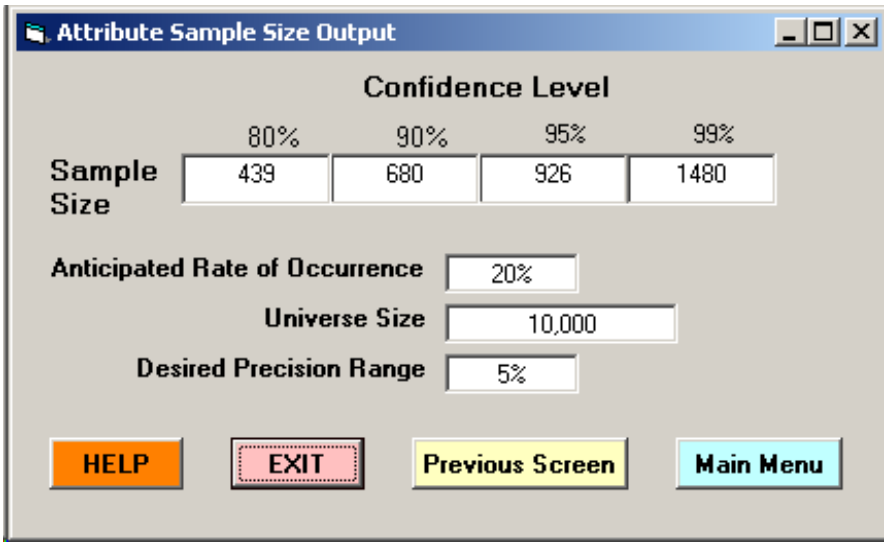
```
NOTE (*):   One or more sample sizes were under 30.  The generated sample
            sizes were the result of mathematical formulas and did not
            incorporate management decisions concerning the purpose of the
            sample or current organizational sampling policies.  You may need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

NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

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.



NOTE: Example is for illustrative purposes only. Be sure to use sample sizes that conform to the organization's minimum sample size standards.

Appendices

Data Limitations for RAT-STATS 2007

Random Number Modules

Module	Limitation
Single Stage	Number of random values (including spares) ≤10,000
Sets of Two	Number of random values (including spares) ≤10,000
Sets of Three	Number of random values (including spares) ≤10,000
Sets of Four	Number of random values (including spares) ≤10,000
Frames - Single Stage	Number of frames in the universe ≤ 50 Number of random values (including spares) ≤10,000
Frames - Sets of Two	Number of frames in the universe ≤ 50 Number of random values (including spares) ≤10,000
RHC Sample Selection	Number of rows in input file ≤ 6,000 Number of units in the sample ≤ 3,000

Attribute Modules

Module	Limitation
Unrestricted	Universe size ≤ 2,147,483,647
Stratified	Number of strata ≤ 125
2-Stage Unrestricted	Number of primary units in the sample ≤ 250
3-Stage Unrestricted	Number of primary units in the sample ≤ 50 Total number of secondary units in the sample ≤ 50,000

RHC 2 Stage	Number of primary units in the sample ≤ 3,000
RHC 3 Stage	Number of primary units in the sample ≤ 50 Total number of secondary units in the sample ≤ 4,000
Stratified Cluster	Number of strata ≤ 100 Number of clusters within each stratum ≤ 500
Stratified Multistage	Number of strata ≤ 500 Universe size for each stratum < 1,000,000,000

Variable Modules

Module	Limitation
Unrestricted	Universe size < 1,000,000,000
Stratified	Number of strata ≤50
2-Stage Unrestricted	Number of primary units in the sample ≤ 500 Number of primary units in the universe < 1,000,000 Number of secondary units in the universe < 1,000,000,000
3-Stage Unrestricted	Number of primary units in the sample ≤ 50 Total number of primary units in the universe ≤ 30,000 Total number of secondary units in the sample ≤ 50,000
RHC 2 Stage	Number of primary units in the sample ≤ 3,000
RHC 3 Stage	Number of primary units in the sample ≤ 50 Total number of secondary units in the sample ≤ 4,000
Stratified Cluster	Number of strata ≤ 100 Number of clusters within each stratum ≤ 500
Stratified Multistage	Number of strata ≤ 500
Poststratified	Number of strata ≤ 150

Unknown Univ. Size	Universe size sample: Universe size < 1,000,000,000 Sample size < 1,000,000 Variable estimation sample: Sample size < 1,000,000

Sample Size Modules

Module	Limitation
Variable Unrestricted (Using a Probe Sample)	Probe sample size must be ≤500,000 (all formats)
Variable Stratified	Number of strata ≤500
Attribute Unrestricted	Anticipated rate of occurrence ≥ .5% Anticipated rate of occurrence ≤ 98% Desired precision range ≥ 1% Desired precision range ≤ 99%

Numerical Accuracy in RAT-STATS 2007

Attribute Modules

Module	Limitation
Unrestricted	Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .05 percent. Calculation of the hypergeometric distribution function is accurate to at least 12 significant figures, provided the universe size is less than 2,147,483,647.
Stratified	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .05 percent.
2-Stage Unrestricted	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .5 percent.
3-Stage Unrestricted	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .5 percent.
RHC 2 Stage	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12.

RHC 3 Stage	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12.
Stratified Cluster	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .5 percent.
Stratified Multistage	Calculations done in this program module are accurate to at least 12 significant digits. Projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Universe proportion point estimates and the associated confidence limits are accurate to the nearest .5 percent.

Variable Modules

Module	Limitation
Unrestricted	Calculations done in this program module are accurate to at least 12 significant digits. The descriptive statistics calculated are accurate to two decimal places, provided the number of significant digits does not exceed 12. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the *t*-distribution are accurate to at least 12 decimal places for degrees of freedom between 1 and 10,000 and confidence levels of 80%, 90%, and 95%.
Stratified	Calculations done in this program module are accurate to at least 12 significant digits. The descriptive statistics calculated are accurate to two decimal places, provided the number of significant digits does not exceed 12. The projected items and the confidence limits for the universe and for each stratum are accurate to the nearest integer, provided the number of significant digits does not exceed 12. Calculation of the critical values of the *t*-distribution are accurate to at least 12 decimal places for degrees of freedom between 1 and 10,000 and confidence levels of 80%, 90%, and 95%.
2-Stage Unrestricted	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.
3-Stage Unrestricted	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.

RHC 2 Stage	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.
RHC 3 Stage	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.
Stratified Cluster	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.
Stratified Multistage	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%. **Note**: The point estimate and confidence limits are accurate to the nearest integer, subject to the limitation that all input values have been rounded to the nearest integer.
Poststratified	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.

Unknown Univ. Size	Calculations done in this program module are accurate to at least 12 significant digits. The projected items in the universe and the universe total confidence limits are accurate to the nearest integer, provided the number of digits does not exceed 12. Calculation of the critical values of the standard normal distribution are accurate to at least 12 decimal places for confidence levels of 80%, 90%, and 95%.

Troubleshooting for RAT-STATS 2007

1. When running RAT-STATS 2007, an error "Error 462 - Printer Error" occurs.

   Solution. Get into the Control Panel and Click on Printers and Faxes. Set one of your available printers as the default printer.

2. When selecting random numbers with output to Excel, the Excel file cannot be viewed.

   Solution. The user must first exit RAT-STATS 2007 before opening the Excel spreadsheet.

3. When clicking on the HELP button, an error message occurs.

   Solution. The User Guide file is already open. Either click on this file in the taskbar to use it or close the User Guide file before clicking on HELP.

4. When opening an Access database file, an error message occurs.

   Solution. The Access file is likely specified as "read only." Locate this file on your computer, right click on it, left click on Properties, and make sure the Read-only box is unchecked.