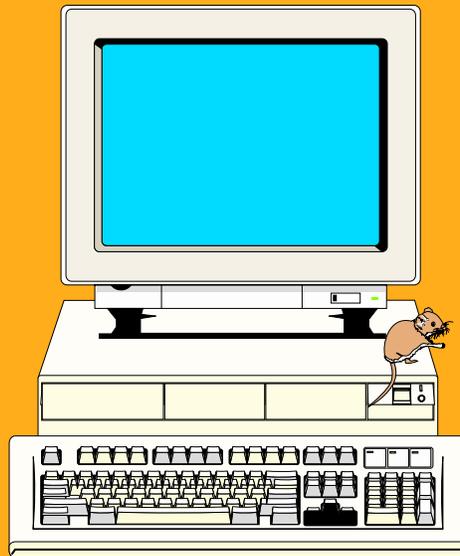


**DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES**

# **RAT-STATS**



# **USER GUIDE**

**Last Update: October 1998 [Y2K Compliant]**

## TABLE OF CONTENTS

	Page
<b>GETTING STARTED</b>	
<b>Introduction</b>	
Introduction .....	i
Installation .....	i
Questions .....	ii
<b>Sample</b>	
Purpose .....	1-1
Random Numbers .....	1-2
Attribute Appraisals .....	1-2
Variable Appraisals .....	1-3
Utility Programs .....	1-4
<b>RANDOM NUMBERS</b>	
<b>Overview</b>	
Overview .....	2-1
<b>Single Stage Numbers</b>	
Purpose .....	2-4
Input Queries .....	2-4
Program Output .....	2-6
Examples of Input Screens .....	2-7
Examples of Program Output .....	2-8
<b>Sets of Two Numbers</b>	
Purpose .....	2-10
Input Queries .....	2-10
Program Output .....	2-12
Examples of Input Screens .....	2-13
Examples of Program Output .....	2-14

**Sets of Three Numbers**

Purpose . . . . .	2-16
Input Queries . . . . .	2-16
Program Output . . . . .	2-18
Examples of Input Screens . . . . .	2-19
Examples of Program Output . . . . .	2-20

**Sets of Four Numbers**

Purpose . . . . .	2-22
Input Queries . . . . .	2-22
Program Output . . . . .	2-24
Examples of Input Screens . . . . .	2-25
Examples of Program Output . . . . .	2-26

**Frames - Single Stage**

Purpose . . . . .	2-28
Input Queries . . . . .	2-28
Program Output . . . . .	2-30
Examples of Input Screens . . . . .	2-31
Examples of Program Output . . . . .	2-32

**Frames - Sets of Two**

Purpose . . . . .	2-34
Input Queries . . . . .	2-34
Program Output . . . . .	2-36
Examples of Input Screens . . . . .	2-37
Examples of Program Output . . . . .	2-39

**RHC Sample Selection**

Purpose . . . . .	2-40
Input Queries . . . . .	2-41
Program Output . . . . .	2-44
Examples of Input Screens . . . . .	2-46
Examples of Program Output . . . . .	2-47

## ATTRIBUTE APPRAISALS

### Overview

Overview .....	3-1
----------------	-----

### Unrestricted

Purpose .....	3-4
Input Queries .....	3-4
Program Output .....	3-6
Example of Input Screen .....	3-7
Example of Program Output .....	3-7

### Stratified

Purpose .....	3-8
Input Queries .....	3-8
Program Output .....	3-11
Example of Input Screen .....	3-12
Example of Program Output .....	3-12

### Two Stage Unrestricted

Purpose .....	3-13
Input Queries .....	3-13
Program Output .....	3-17
Example of Input Screen .....	3-18
Example of Program Output .....	3-19

### Three Stage Unrestricted

Purpose .....	3-20
Input Queries .....	3-20
Program Output .....	3-23
Example of Input Screens .....	3-25
Example of Program Output .....	3-26

### Stratified Cluster

Purpose .....	3-28
Input Queries .....	3-28
Program Output .....	3-30
Example of Input Screen .....	3-33
Example of Program Output .....	3-33

**Stratified Multistage**

Purpose .....	3-34
Input Queries .....	3-34
Program Output .....	3-37
Example of Input Screen .....	3-38
Example of Program Output .....	3-38

**VARIABLE APPRAISALS****Overview**

Overview .....	4-1
----------------	-----

**Unrestricted**

Purpose .....	4-5
Input Queries .....	4-5
Program Output .....	4-9
Example of Input Screen .....	4-11
Example of Program Output .....	4-12

**Stratified**

Purpose .....	4-13
Input Queries .....	4-13
Program Output .....	4-18
Example of Input Screens .....	4-21
Example of Program Output .....	4-22

**Two-Stage Unrestricted**

Purpose .....	4-23
Input Queries .....	4-23
Program Output .....	4-28
Example of Input Screens .....	4-30
Example of Program Output .....	4-32

**Three-Stage Unrestricted**

Purpose .....	4-33
Input Queries .....	4-33
Program Output .....	4-37
Example of Input Screen .....	4-39
Example of Program Output .....	4-40

---

<b>Stratified Cluster</b>	
Purpose .....	4-41
Input Queries .....	4-41
Program Output .....	4-44
Example of Input Screen .....	4-48
Example of Program Output .....	4-48
<b>Stratified Multistage</b>	
Purpose .....	4-49
Input Queries .....	4-49
Program Output .....	4-52
Example of Input Screen .....	4-53
Example of Program Output .....	4-53
<b>RHC Two-Stage</b>	
Purpose .....	4-54
Input Queries .....	4-54
Program Output .....	4-59
Example of Input Screen .....	4-62
Example of Program Output .....	4-63
<b>RHC Three-Stage</b>	
Purpose .....	4-64
Input Queries .....	4-64
Program Output .....	4-69
Example of Input Screen .....	4-72
Example of Program Output .....	4-72
<b>Post Stratification</b>	
Purpose .....	4-75
Input Queries .....	4-75
Program Output .....	4-80
Example of Input Screen .....	4-82
Example of Program Output .....	4-83
<b>Unknown Universe</b>	
Purpose .....	4-84
Input Queries .....	4-84
Program Output .....	4-86
Example of Input Screens .....	4-88
Example of Program Output .....	4-89

---

---

---

## UTILITY PROGRAMS

### Overview

Overview .....	5-1
----------------	-----

### File Dump Utility

Purpose .....	5-3
Input Queries .....	5-3
Program Output .....	5-4
Example of Input Screen .....	5-5
Example of Program Output .....	5-5

### Sample Size Estimators

Purpose .....	5-6
Input Queries .....	5-6
Program Output .....	5-8
Example of Input Screen .....	5-10
Example of Program Output .....	5-10

### Add Decimal Points

Purpose .....	5-11
Input Queries .....	5-11
Program Output .....	5-13
Example of Input Screen .....	5-14
Example of Program Output .....	5-15

### Transposition Error Identification

Purpose .....	5-16
Input Queries .....	5-16
Program Output .....	5-17
Example of Input Screen .....	5-18
Example of Program Output .....	5-18

## INTRODUCTION

**DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RAT-STATS  
VERSION: MAY 1993**

**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 Disk Operating System (MS-DOS).

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 **RAT-STATS** package has been compressed using PKWARE Version 1.1 software. The self-extracting file is stored as **RATSTATS.EXE**. The PKUNZIP extraction process is performed by first setting the DOS prompt to the drive and directory in which **RATSTATS.EXE** is stored. After determining where the RAT-STATS package is to be stored after being extracted (the package requires about 560k of space), a command is entered using the following format:

**RATSTATS C:\PATH**

Explanation:

**C:\** is the drive where the package will be stored.

**PATH** is the path to the directory on the drive where the package will be stored.

This package also allows the user to customize the heading for query screens and output. A maximum of two lines may be created. Typically, the name of the organization and location are placed on the two lines.

To execute the program, the user must be in the directory that the RAT-STATS package. Once in the directory, the user types **RATSET** and hits the Enter key. The program will prompt for the two line heading. Each line is limited to 60 positions. The user only needs to execute this program when the two line heading needs to be changed.

To use RAT-STATS, after installation, enter the extracted disk into the appropriate drive (or select the appropriate directory if the programs have been installed onto a hard disk), and type the word **SAMPLE** followed by the Enter key.

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.

## Questions

Since this is the third update of the documentation, we decided to take a little time to answer some frequently asked questions about the software.

### **ARE THE RANDOM NUMBERS REALLY RANDOM?**

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

### **WHAT IS THE COMPUTER LANGUAGE OF RAT-STATS?**

The software was written in Turbo BASIC (version 1.1). The BASIC compiler was developed by Borland International.

## **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, University of North Texas  
Doug Rennie - Manager, RATS - San Francisco

## **WHO DO I CALL ABOUT RAT-STATS?**

If you have questions about the software or suggestions for improvements, you may contact Doug Rennie on the Internet at [drennie@os.dhhs.gov](mailto:drennie@os.dhhs.gov).

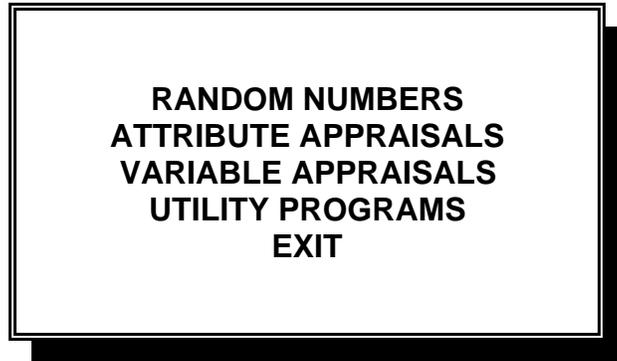
## **WHY IS IT CALLED RAT-STATS?**

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

## **WILL THERE BE FUTURE UPDATES?**

Based on the responses to earlier versions of the package, we will probably continue to update the package based on the request of users.

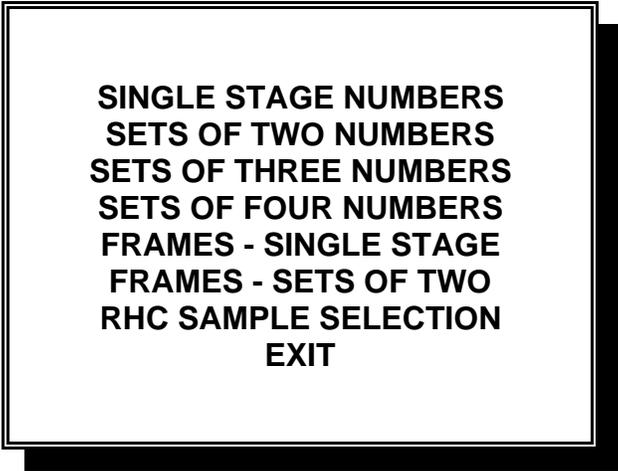
## SAMPLE



### Purpose

**SAMPLE** is the host program for the RAT-STATS statistical sampling package. The primary menu offers the user four categories of applications. By using the cursor and enter keys, the user will be able to select a particular category. Each category has its own set of selection options.

## 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  
EXIT**

The RAT-STATS package has four categories of applications. The first category is a series of random number generators that should be used to randomly select items for review. By highlighting **RANDOM NUMBERS** and then hitting the Enter key, the user will see a new window appear on the screen with a variety of random number programs. Section two of this guide explains how to use each of the random number programs.

## ATTRIBUTE APPRAISALS



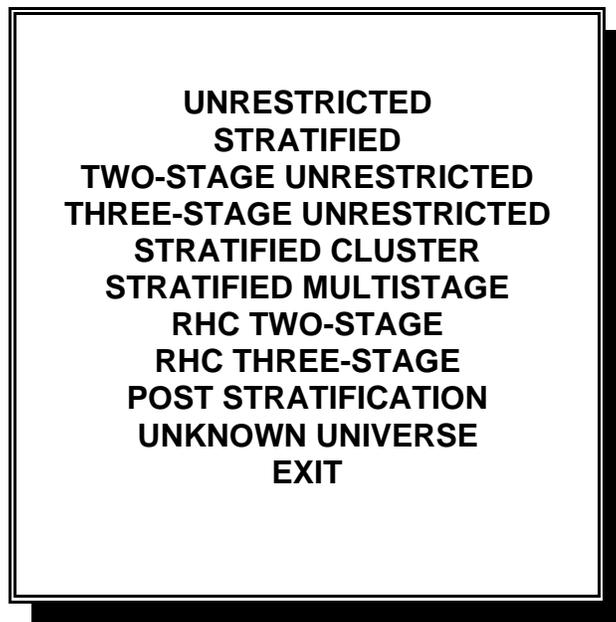
**UNRESTRICTED  
STRATIFIED  
TWO-STAGE UNRESTRICTED  
THREE-STAGE UNRESTRICTED  
STRATIFIED CLUSTER  
STRATIFIED MULTISTAGE  
EXIT**

Attribute appraisals allows a user to estimate the rate of occurrence of a given condition. The user may, for example, want to know the rate of occurrence of checks issued for less

than \$100. After drawing a random sample and evaluating the items selected, an attribute appraisal would be used not only to estimate the rate of occurrence, but also to determine (with a measured degree of confidence) the boundaries of the estimate.

By highlighting **ATTRIBUTE APPRAISALS** and hitting the Enter key, the user will see a window appear on the screen with a variety of attribute appraisal programs. Section three of this guide explains each of the attribute appraisal programs.

### VARIABLE APPRAISALS

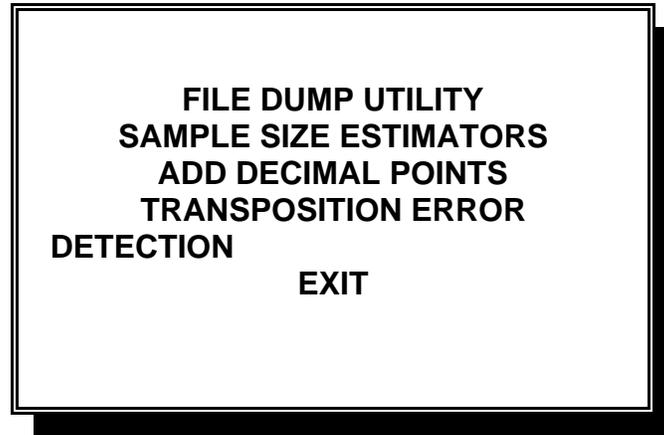


The purpose of using variable appraisals is to measure a quantitative characteristic or set of characteristics. The user may, for example, want to know the value of all checks approved by a certain supervisor. After drawing a random sample and identifying the checks approved by the supervisor, a variable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

By highlighting **VARIABLE APPRAISALS** and hitting the Enter key, the user will see a window appear on the screen with a variety of variable appraisal programs. Section four of this guide explains each of the variable appraisal programs.

---

## UTILITY PROGRAMS

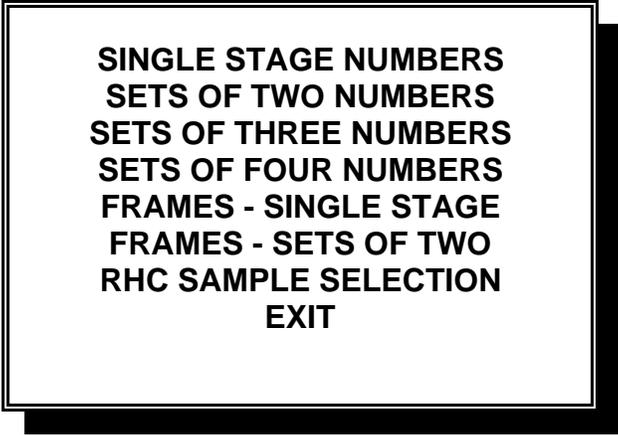


To assist the user in working with data used in statistical applications, some utility programs have been included in this package. By highlighting **UTILITY PROGRAMS** and hitting the Enter key, the user will see a window appear on the screen with a series of program titles. Section five of this guide explains each of the utility programs.

### EXIT

The **EXIT** option will allow the user to leave that level of menu options and return to the primary menu. From the primary menu, the **EXIT** option will cause the user to leave the RAT-STATS package.

## 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  
EXIT**

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 universe 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 universe may not make it feasible to manually create such a sequencing technique.

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

This package attempts to minimize the efforts of the user in actually identifying the items to be sampled. Once the user has identified the boundaries of the universe, the random number module will ensure that only an unduplicated list of random numbers within the stated range are selected. Each module also allows the user to have all or a portion of the random numbers to be sorted in ascending order to allow for a more efficient retrieval of the sampled items. The user also has the option of having the random numbers printed or sent to a disk file.

The user of this package should determine the most efficient manner of selecting

sample items and then use the appropriate module from the list above to generate the random numbers. A brief description of when to use each module is given below. A detailed explanation of how to use each module is described later in this section.

### **SINGLE STAGE NUMBERS**

This module is used when the universe of items has a numbering scheme. The universe could be a computer listing with each item numbered from one to the size of the universe. 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 one 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 1201 through 1483 and 2833 through 3261. The module allows the user to

enter in these two range sets (frames). The module will calculate the universe (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. 1210).

The user may also have a situation where vouchers are numbered starting with 1 at the beginning of each month. For a six 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 NUMBERS**

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 universe of items.

### Input Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

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.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers to be generated in sequential order may be entered with or without commas. After the quantity indicated has been generated by the program, the random numbers will be sorted and the output will be in ascending order to assist the user in retrieving the sample items. The order of selection will be printed with the random numbers. If the quantity desired is zero, then a "0" (zero) must be entered.

### SPARES IN RANDOM ORDER?

The quantity of numbers to be generated in random order may be entered with or without commas. The random numbers will be displayed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

### ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS):

The low and high numbers in the universe are the boundaries of the population from which the user will be sampling. If the universe is a computer listing numbered 1 through 1,345 then the low entry will be 1 and the high entry will be 1,345 (commas may be used). If the universe 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.

### DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER?

The random numbers may be sent to the printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

### ENTER NAME FOR FILE TO RECORD THE RESULTS?

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive could be "**A:FILEONE.NBR**" or "**A:\AUDIT\OUTFILE1.NBR**" if it was going to a directory on the "A" drive.

### ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Two examples of the query screen appear on page 2-7. The user entries are in **boldface**. Pages 2-8 and 2-9 have examples of the output going to the printer and to a disk file based on the responses shown in the query examples.

## Program Output

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.

The random numbers in generated order will begin on a new page after all of the numbers in sequential order have been printed. At the bottom of the output will appear a check figure. This is the summation of all the random numbers generated for output. This sum can be used to verify that the numbers are transcribed correctly to another file or worksheet.

If the user selects the disk for the output, then the summary information will be appended to the file as well as displayed on the screen.

DATE: 07/23/92	DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR VERSION: MONTH YEAR	TIME: 14:22
----------------	---	-------------

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
ELSE HIT THE ENTER KEY?

ENTER THE NAME OF THE AUDIT/REVIEW? EXAMPLE OF OUTPUT GOING TO THE PRINTER

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
SEQUENTIAL ORDER? **10** SPARES IN RANDOM ORDER? **5**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS):  
LOW: ? **1**  
HIGH: ? **1,000**

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

DATE: 07/23/92	DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER PROGRAM VERSION: MONTH YEAR	TIME: 14:22
----------------	---	-------------

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
ELSE HIT THE ENTER KEY? **147682**

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF OUTPUT TO DISK**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
SEQUENTIAL ORDER? **1831** SPARES IN RANDOM ORDER? **0**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS):  
LOW: ? **6,435**  
HIGH: ? **94,736**

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **D**

ENTER NAME OF FILE TO RECORD THE RESULTS? **A:\AUDIT\OUTFILE1.NBR**

DEPARTMENT OF HEALTH & HUMAN SERVICES					
OIG - OFFICE OF AUDIT SERVICES					
RANDOM NUMBER GENERATOR					
DATE: 07/23/92					TIME: 14:22
AUDIT: EXAMPLE OF OUTPUT GOING TO THE PRINTER					
SEED NUMBER: 28703.84305469418			UNIVERSE SIZE: 1,000		
10 RANDOM NUMBERS IN SEQUENTIAL ORDER					
ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER
7 -	94	8 -	393	3 -	495
4 -	511	2 -	596	6 -	650
10 -	663	1 -	711	5 -	884
9 -	978				
=====NEW PAGE=====					
DEPARTMENT OF HEALTH & HUMAN SERVICES					
OIG - OFFICE OF AUDIT SERVICES					
RANDOM NUMBER GENERATOR					
DATE: 07/23/92					TIME: 14:22
AUDIT: EXAMPLE OF OUTPUT GOING TO THE PRINTER					
SEED NUMBER: 28703.84305469418			UNIVERSE SIZE: 1,000		
5 RANDOM NUMBERS IN GENERATED ORDER					
ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER	ORDER OF SELECTION	RANDOM NUMBER
11 -	37	12 -	996	13 -	183
14 -	254	15 -	787		
SUMMATION OF RANDOM NUMBERS 8,232					

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER GENERATOR

DATE: 07/23/92 TIME: 14:22

AUDIT: EXAMPLE OF OUTPUT TO DISK

SEED NUMBER: 147682 UNIVERSE SIZE: 88,302

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

TOTAL RANDOM NUMBERS GENERATED: 1,831

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:  
POSITIONS 1 THROUGH 6 - ORDER OF SELECTION  
POSITIONS 7 THROUGH 17 - RANDOM NUMBER  
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.  
SUMMATION OF RANDOM NUMBERS = 92,465,523

---

---

## 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 Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

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.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers generated in sequential order may be entered with or without commas. 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 a "0" (zero) must be entered.

**SPARES IN RANDOM ORDER?**

The quantity of numbers to be generated in random order may be entered with or

without commas. The random numbers will be printed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

**ENTER THE UNIVERSE RANGE (LOW & HIGH) FOR:**

	<b>FIRST SET</b>	<b>SECOND SET</b>
<b>LOW:</b>	?	?
<b>HIGH:</b>	?	?

The program prompts the user to enter the low and high numbers of the universe that constitute the first range in the pair of ranges for the entire universe. 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 universe 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 universe 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.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR  
<P>RINTER?**

The random numbers may be sent to a printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

**ENTER NAME FOR FILE TO RECORD THE  
RESULTS?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive could be "**A:FILEONE.NBR**" or "**A:\AUDIT\OUTFILE1.NBR**" if it were going to a directory on the "A" drive.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that

use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Examples of the query screen appear on page 2-13. The user entries are in **boldface**. The related output is shown on pages 2-14 and 2-15.

## Program Output

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 universe 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.

If the user selects the disk for the output, then only summary information will be sent to the printer. If the printer is not connected to the computer, then the summary information will be displayed on the screen.

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 06/01/93	RANDOM NUMBER PROGRAM	TIME: 14:22
VERSION: MONTH YEAR		
IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY?		
ENTER THE NAME OF THE AUDIT/REVIEW? <b>EXAMPLE OF PAIRS GOING TO PRINT</b>		
ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN SEQUENTIAL ORDER? <b>10</b> SPARES IN RANDOM ORDER? <b>4</b>		
ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:		
	FIRST SET	SECOND SET
LOW:	? <b>1</b>	? <b>1</b>
HIGH:	? <b>658</b>	? <b>66</b>
DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? <b>P</b>		
ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)		

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 06/01/93	RANDOM NUMBER PROGRAM	TIME: 14:22
VERSION: MONTH YEAR		
IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY? <b>147682</b>		
ENTER THE NAME OF THE AUDIT/REVIEW? <b>EXAMPLE OF PAIRS GOING TO DISK</b>		
ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN SEQUENTIAL ORDER? <b>1,831</b> SPARES IN RANDOM ORDER? <b>0</b>		
ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:		
	FIRST SET	SECOND SET
LOW:	? <b>1</b>	? <b>1</b>
HIGH:	? <b>658</b>	? <b>66</b>
DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? <b>D</b>		
ENTER NAME FOR FILE TO RECORD THE RESULTS? <b>A:\AUDIT\OUTFILE1.NBR</b>		

DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR			TIME: 14:22	
AUDIT: EXAMPLE OF PAIRS GOING TO PRINT						
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 43,428				
10 RANDOM NUMBERS IN SEQUENTIAL ORDER						
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	
7 -	10	46	5 -	65	45	
9 -	122	1	10 -	336	29	
1 -	475	8	8 -	482	50	
2 -	520	10	6 -	630	59	
4 -	635	46	3 -	641	61	
=====NEW PAGE=====						
DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR			TIME: 14:22	
AUDIT: EXAMPLE OF PAIRS GOING TO PRINT						
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 43,428				
4 RANDOM NUMBERS IN GENERATED ORDER						
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	
11 -	161	43	12 -	413	30	
13 -	197	40	14 -	648	15	

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER GENERATOR

DATE: 06/01/93 TIME: 14:22

AUDIT: EXAMPLE OF PAIRS GOING TO DISK

SEED NUMBER: 147682                      UNIVERSE SIZE: 43,428

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

TOTAL RANDOM NUMBERS GENERATED: 1,831

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 28 - SECOND NUMBER OF SET  
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.

---

---

## 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 Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

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.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers to be generated in sequential order may be entered with or without commas. 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 a "0" (zero) must be entered.

**SPARES IN RANDOM ORDER?**

The quantity of numbers to be generated in random order may be entered with or

without commas. The random numbers will be printed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

**ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS):**

	<b>FIRST SET</b>	<b>SECOND SET</b>	<b>THIRD SET</b>
<b>LOW:</b>	?	?	?
<b>HIGH:</b>	?	?	?

The low and high numbers in the universe are the boundaries of the population 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 659 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 659. The low for the third set would be 1 and the high would be 66. The overall universe for this sample would be 521,928 (12 times 659 times 66). For ease of use the path to the sample items should dictate the order of universe 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.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER?**

The random numbers may be sent to a printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

**ENTER NAME FOR FILE TO RECORD THE RESULTS?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive would be "**A:FILEONE.NBR**" or "**A:\AUDIT\OUTFILE1.NBR**" if it were going to a directory on the "A" drive.

## ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Examples of the query screen appear on page 2-19. The user entries are in **boldface**. The related output is shown on pages 2-20 and 2-21.

### Program Output

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 universe 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. For numbers printed in sequential order, the lowest set of numbers is printed first. The subsequent sets of random numbers are in sequential order with one set of numbers on each line.

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

If the user selects the disk for the output, then only summary information will be sent to the printer. If the printer is not connected to the computer, then the summary information will be displayed on the screen.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER PROGRAM  
VERSION: MONTH YEAR

DATE: 06/01/93 TIME: 14:22

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
ELSE HIT THE ENTER KEY?

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SETS OF THREE**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
SEQUENTIAL ORDER? **4** SPARES IN RANDOM ORDER? **2**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:  
FIRST SET SECOND SET THIRD SET

LOW:	? <b>1</b>	? <b>1</b>	? <b>1</b>
HIGH:	? <b>12</b>	? <b>659</b>	? <b>66</b>

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER PROGRAM  
VERSION: MONTH YEAR

DATE: 06/01/93 TIME: 14:22

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
ELSE HIT THE ENTER KEY? **147682**

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SETS OF THREE**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
SEQUENTIAL ORDER? **1,831** SPARES IN RANDOM ORDER? **0**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:  
FIRST SET SECOND SET THIRD SET

LOW:	? <b>1</b>	? <b>1</b>	? <b>1</b>
HIGH:	? <b>12</b>	? <b>659</b>	? <b>66</b>

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **D**

ENTER NAME FOR FILE TO RECORD THE RESULTS? **A:IAUDITOUTFILE1.NBR**

DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR		TIME: 14:23
AUDIT: EXAMPLE OF SETS OF THREE				
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 521,928		
4 RANDOM NUMBERS IN SEQUENTIAL ORDER				
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	
4 -	4	8	26	
1 -	6	54	53	
3 -	11	545	2	
2 -	12	486	40	
=====NEW PAGE=====				
DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR		TIME: 14:23
AUDIT: EXAMPLE OF SETS OF THREE				
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 521,928		
2 RANDOM NUMBERS IN GENERATED ORDER				
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	
5 -	3	65	38	
6 -	10	45	37	

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER GENERATOR

DATE: 06/01/93

AUDIT: EXAMPLE OF SETS OF THREE

SEED NUMBER: 147682                      UNIVERSE SIZE: 521,928

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

TOTAL RANDOM NUMBERS GENERATED: 1,831

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 28 - SECOND NUMBER OF SET  
POSITIONS 29 THROUGH 39 - THIRD NUMBER OF SET  
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.

---

---

## 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 Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

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.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers generated in sequential order may be entered with or without commas. 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 a "0" (zero) must be entered.

**SPARES IN RANDOM ORDER?**

The quantity of numbers to be generated in random order may be entered with or

without commas. The random numbers will be printed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

**ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS):**

	FIRST SET	SECOND SET	THIRD SET	FOURTH SET
<b>LOW:</b>	?	?	?	?
<b>HIGH:</b>	?	?	?	?

The program prompts the user to enter the low and high numbers for each of the four sets of universes. If, for example, the user was planning to select items from five 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 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 universe 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 universe 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 the 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.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER?**

The random numbers may be sent to the printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

**ENTER NAME FOR FILE TO RECORD THE RESULTS?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive could be "**A:FILEONE.NBR**" or "**A:\AUDIT\OUTFILE1.NBR**" if it were going to a directory on the "A" drive.

## ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Examples of the query screen appear on page 2-25. The user entries are in **boldface**. The related output is shown on pages 2-26 and 2-27.

### Program Output

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 universe boundaries. The third piece of information is the random number generated from the second set of boundaries. The fourth piece of information is the number generated from the third set of boundaries. The fifth piece of information is the number generated from the fourth 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 with one set of numbers on each line.

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

If the user selects the disk for the output, then only summary information will be sent to the printer. If the printer is not connected to the computer, then the summary information will be displayed on the screen.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 RANDOM NUMBER PROGRAM  
 VERSION: MONTH YEAR

DATE: 06/01/90 TIME: 14:22

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
 ELSE HIT THE ENTER KEY?

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SETS OF FOUR**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
 SEQUENTIAL ORDER? **3** SPARES IN RANDOM ORDER? **2**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:

	FIRST SET	SECOND SET	THIRD SET	FOURTH SET
LOW:	? 1	? 1	? 1	? 1
HIGH:	? 5	? 12	? 658	? 66

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 RANDOM NUMBER PROGRAM  
 VERSION: MONTH YEAR

DATE: 06/01/90 TIME: 14:22

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
 ELSE HIT THE ENTER KEY? **147682**

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SETS OF FOUR**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
 SEQUENTIAL ORDER? **1,831** SPARES IN RANDOM ORDER? **0**

ENTER THE UNIVERSE RANGE (LOW & HIGH NUMBERS) FOR:

	FIRST SET	SECOND SET	THIRD SET	FOURTH SET
LOW:	? 1	? 1	? 1	? 1
HIGH:	? 5	? 12	? 658	? 66

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? **D**

ENTER NAME FOR FILE TO RECORD THE RESULTS? **A:AUDITOUTFILE1.NBR**

DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR				
DATE: 06/01/93				TIME: 14:23
AUDIT: EXAMPLE OF SETS OF FOUR				
SEED NUMBER: 28703.84305469418				UNIVERSE SIZE: 2,605,680
3 RANDOM NUMBERS IN SEQUENTIAL ORDER				
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	FOURTH NBR.
2 -	2	3	49	55
1 -	3	9	465	53
3 -	5	12	346	19
=====NEW PAGE=====				
DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR				
DATE: 06/01/90				TIME: 14:23
AUDIT: EXAMPLE OF SETS OF FOUR				
SEED NUMBER: 28703.84305469418				UNIVERSE SIZE: 2,605,680
2 RANDOM NUMBERS IN GENERATED ORDER				
ORDER OF SELECTION	FIRST NBR.	SECOND NBR.	THIRD NBR.	FOURTH NBR.
4 -	2	7	290	29
5 -	3	7	374	8

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER GENERATOR

DATE: 06/01/93

AUDIT: EXAMPLE OF SETS OF FOUR

SEED NUMBER: 147682                      UNIVERSE SIZE: 2,605,680

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

TOTAL RANDOM NUMBERS GENERATED: 1,831

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 28 - SECOND NUMBER OF SET  
POSITIONS 29 THROUGH 39 - THIRD NUMBER OF SET  
POSITIONS 40 THROUGH 50 - FOURTH NUMBER IN SET  
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.

---

## 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 1050 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 re-number the subuniverse of desired transactions, this program could be used with the existing numbering system.

### Input Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

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 (such as from a book of random numbers). The seed number will be printed as part of the output.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers to be generated in sequential order may be entered with or without commas. 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 a "0" (zero) must be entered.

### **SPARES IN RANDOM ORDER?**

The quantity of numbers to be generated in random order may be entered with or without commas. The random numbers will be printed in the order selected. If the quantity desired is zero, then a "0" must 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 example described above, the user would have two frames. The first would contain items 1 through 1,050 and the second frame would contain items 8,405 through 9,565.

### **FRAME LOW NBR HIGH NBR    FRAME LOW NBR HIGH NBR**

Based on the frame count entered by the user, the program will prompt the user for the low and high boundaries for each frame. Each entry may include commas in the number.

### **DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER?**

The random numbers may be sent to the printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

### **ENTER NAME FOR FILE TO RECORD THE RESULTS?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive would be "**A:FILEONE.NBR**".

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only

hitting the Enter key.

Two examples of the query screen appear on page 2-31. The user entries are in **boldface**. Pages 2-32 and 2-33 have examples of the output going to the printer or to a disk file based on the responses shown in the query examples.

### Program Output

For each random number 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. The next two pieces of information result in converting the random number into the appropriate frame and the item within the frame. For numbers printed in sequential order of the printout, the lowest random number and the corresponding conversion is printed first. The subsequent numbers are in sequential order with one 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.

If the user selects the disk for the output, then only summary information will be sent to the printer. If a printer is not connected to the computer, then the summary information will be displayed on the screen.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 RANDOM NUMBER PROGRAM  
 VERSION: MONTH YEAR

DATE: 06/01/93 TIME: 14:23

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
 ELSE HIT THE ENTER KEY?

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SINGLE FRAMES**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
 SEQUENTIAL ORDER? 3      SPARES IN RANDOM ORDER? 2

ENTER THE NUMBER OF FRAMES IN YOUR UNIVERSE (MAX 51)? 2

FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR	
=====	=====	=====	=====	=====	=====	=====	=====	=====	
1	?	1	?	1,050	2	?	8,405	?	9,565

DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 RANDOM NUMBER PROGRAM  
 VERSION: MONTH YEAR

DATE: 06/01/93 TIME: 14:22

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
 ELSE HIT THE ENTER KEY? **147682**

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SINGLE FRAMES**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
 SEQUENTIAL ORDER? 5      SPARES IN RANDOM ORDER? 0

ENTER THE NUMBER OF FRAMES IN YOUR UNIVERSE (MAX 51)? 2

FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR	
=====	=====	=====	=====	=====	=====	=====	=====	=====	
1	?	1	?	1,050	2	?	8,405	?	9,565

OUTPUT RANDOM NUMBERS TO <D>ISK OR <P>RINTER? D

ENTER OUTPUT FILE NAME FOR THE RANDOM NUMBERS? **A:\AUDIT\OUTFILE1.NBR**

DEPARTMENT OF HEALTH & HUMAN SERVICES			
OIG - OFFICE OF AUDIT SERVICES			
DATE: 06/01/93	RANDOM NUMBER GENERATOR		TIME: 14:23
AUDIT: EXAMPLE OF SINGLE FRAMES			
SEED NUMBER: 53152.44337317086		UNIVERSE SIZE: 2,211	
3 RANDOM NUMBERS IN SEQUENTIAL ORDER			
ORDER OF SELECTION	RANDOM NUMBER	-CONVERTED NUMBER-FRAME	ITEM NUMBER
1 -	368	1	368
2 -	850	1	850
3 -	1,340	2	8,694
=====NEW PAGE=====			
DEPARTMENT OF HEALTH & HUMAN SERVICES			
OIG - OFFICE OF AUDIT SERVICES			
DATE: 06/01/93	RANDOM NUMBER GENERATOR		TIME: 14:23
AUDIT: EXAMPLE OF SINGLE FRAMES			
SEED NUMBER: 53152.44337317086		UNIVERSE SIZE: 2,211	
2 RANDOM NUMBERS IN GENERATED ORDER			
ORDER OF SELECTION	RANDOM NUMBER	-CONVERTED NUMBER-FRAME	ITEM NUMBER
4 -	98	1	98
5 -	1,958	2	9,312

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
RANDOM NUMBER GENERATOR

DATE: 06/01/93 TIME: 14:22

AUDIT: EXAMPLE OF SINGLE FRAMES

SEED NUMBER: 147682 UNIVERSE SIZE: 2,211

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

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 - RANDOM NUMBER  
POSITIONS 18 THROUGH 19 - FRAME NUMBER  
POSITIONS 20 THROUGH 30 - ITEM NUMBER WITHIN FRAME  
EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED.

---

## 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 a computer printout that has page and line numbers. The transactions of interest may be on pages numbered from 1 through 1,050 and begin again at page 8,405 and run through 9,565. Rather than selecting random numbers between 1 and 9,565 and having to discard numbers between 1,050 and 8,405 or have the user re-number the pages of desired transactions, this program could be used with the existing numbering system.

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 Queries

**IF YOU WANT TO PROVIDE A RANDOM START,  
ENTER THE SEED NUMBER; ELSE HIT THE ENTER  
KEY?**

The program allows the 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.

**ENTER THE NAME OF THE AUDIT/REVIEW?**

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

---

---

**ENTER THE QUANTITY OF NUMBERS DESIRED TO  
BE GENERATED IN SEQUENTIAL ORDER?**

The quantity of random numbers generated in sequential order may be entered with or without commas. 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 a "0" (zero) must be entered.

**SPARES IN RANDOM ORDER?**

The quantity of numbers to be generated in random order may be entered with or without commas. The random numbers will be printed in the order selected. If the quantity desired is zero, then a "0" (zero) must be entered.

**ENTER THE NUMBER OF FRAMES IN YOUR  
UNIVERSE (MAX 51)?**

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

**FRAME LOW NBR HIGH NBR FRAME LOW NBR HIGH NBR**

Based on the frame count entered by the user, the program will prompt the user for the low and high boundaries for each primary frame. Each entry may include commas.

**WILL THE RANGE WITHIN EACH FRAME BE THE  
SAME (YES OR NO)?**

If the range within each frame is the same, then the user will only be prompted once for the low and high boundaries. For example, if there are 50 items within each frame, then the response to this query would be **YES**.

**WHAT IS THE LOWEST NUMBER IN EACH SUBFRAME?  
WHAT IS THE HIGHEST NUMBER IN EACH SUBFRAME?**

If the range is the same then the user only needs to enter the low and high numbers once. However, if the range varies from frame to frame, then the user will be prompted to enter the low and high numbers of each frame. Commas may be included with each entry.

### DO YOU WANT THE RESULTS PRINTED TO <D>ISK OR <P>RINTER?

The random numbers may be sent to the printer or to a disk file. The user selects the appropriate output by responding to the query with either a **D** (disk) or a **P** (printer).

### ENTER NAME FOR FILE TO RECORD THE RESULTS?

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example a disk file going to the "A" drive could be "**A:FILEONE.NBR**" or "**A:\AUDIT\OUTFILE1.NBR**" if it were going to a directory on the "A" drive.

### ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Two examples of the query screen are on pages 2-38 and 2-39. The user entries are in **boldface**. Pages 2-40 and 2-41 have examples of the output.

### Program Output

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 result from converting the random number into the appropriate frame, sub-frame and the item within the sub-frame. In the sequential portion of the printout, the lowest random number and the corresponding conversion is printed first. The subsequent numbers are in sequential order with one 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.

If the user selects the disk for the output, then the summary information will be appended to the file as well as displayed on the screen.

```

                                DEPARTMENT OF HEALTH & HUMAN SERVICES
                                OIG - OFFICE OF AUDIT SERVICES
DATE: 06/01/93                                RANDOM NUMBER PROGRAM                                TIME: 14:23
                                VERSION: MONTH YEAR

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;
ELSE HIT THE ENTER KEY?

ENTER THE NAME OF THE AUDIT/REVIEW? EXAMPLE OF SETS OF TWO FRAMES

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN
SEQUENTIAL ORDER? 3    SPARES IN RANDOM ORDER? 2

ENTER THE NUMBER OF FRAMES IN YOUR UNIVERSE (MAX 51)? 3

FRAME  LOW NBR  HIGH NBR  FRAME  LOW NBR  HIGH NBR  FRAME  LOW NBR  HIGH NBR
=====  =====  =====  =====  =====  =====  =====  =====  =====
  1    ?      1    ?    100    2    ?      1    ?    456    3    ?      45
?                               ?                               ?                               ?
                               832

                                -----NEW SCREEN-----

WILL THE RANGE WITHIN EACH FRAME BE THE SAME (YES OR NO)? Y

WHAT IS THE LOWEST NUMBER IN EACH SUBFRAME? 1

WHAT IS THE HIGHEST NUMBER IN EACH SUBFRAME? 66

OUTPUT RANDOM NUMBERS TO <D>ISK OR <P>RINTER? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?
    
```

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 RANDOM NUMBER PROGRAM  
 VERSION: MONTH YEAR

DATE: 06/01/93 TIME: 14:23

IF YOU WANT TO PROVIDE A RANDOM START, ENTER THE SEED NUMBER;  
 ELSE HIT THE ENTER KEY? **147682**

ENTER THE NAME OF THE AUDIT/REVIEW? **EXAMPLE OF SETS OF TWO FRAMES**

ENTER THE QUANTITY OF NUMBERS DESIRED TO BE GENERATED IN  
 SEQUENTIAL ORDER? **831** SPARES IN RANDOM ORDER? **0**

ENTER THE NUMBER OF FRAMES IN YOUR UNIVERSE (MAX 51)? **3**

FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR	FRAME	LOW NBR	HIGH NBR
1	?	1	2	?	1	3	?	45
?		<b>832</b>			<b>456</b>			

-----NEW SCREEN-----

WILL THE RANGE WITHIN EACH FRAME BE THE SAME (YES OR NO)? **N**

-----NEW SCREEN-----

FRAME	LOW NUMBER	HIGH NUMBER	LOW NUMBER	HIGH NUMBER
1	1	100	?	1
?	<b>66</b>			
2	1	456	?	1
?	<b>50</b>			
3	45	832	?	1
			?	<b>66</b>

OUTPUT RANDOM NUMBERS TO <D>ISK OR <P>RINTER? **D**

ENTER OUTPUT FILE NAME FOR THE RANDOM NUMBERS? **A:\AUDIT\OUTFILE1.NBR**

DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR			TIME: 14:23
AUDIT: EXAMPLE OF SETS OF TWO FRAMES					
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 88,704			
3 RANDOM NUMBERS IN SEQUENTIAL ORDER					
ORDER OF SELECTION	RANDOM NUMBER	--- CONVERTED NUMBER ---			
		FRAME	SUB-FRAME	ITEM NUMBER	
1 -	29,297	2	344	59	
2 -	37,787	3	61	35	
3 -	80,711	3	711	59	
=====NEW PAGE=====					
DATE: 06/01/93		DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES RANDOM NUMBER GENERATOR			TIME: 14:23
AUDIT: EXAMPLE OF SETS OF TWO FRAMES					
SEED NUMBER: 28703.84305469418		UNIVERSE SIZE: 88,704			
2 RANDOM NUMBERS IN GENERATED ORDER					
ORDER OF SELECTION	RANDOM NUMBER	--- CONVERTED NUMBER ---			
		FRAME	SUB-FRAME	ITEM NUMBER	
4 -	34,889	2	429	41	
5 -	62,561	3	436	59	

```
DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 06/01/93          RANDOM NUMBER GENERATOR          TIME: 14:23

AUDIT: EXAMPLE OF SETS OF TWO FRAMES

SEED NUMBER: 147682          UNIVERSE SIZE: 81,408

FILE OF RANDOM NUMBERS: A:\AUDIT\OUTFILE1.NBR

TOTAL RANDOM NUMBERS GENERATED: 831

THE NUMBERS ARE IN THE FOLLOWING FORMAT IN YOUR FILE:
    POSITIONS 1 THROUGH 6 - ORDER OF SELECTION
    POSITIONS 7 THROUGH 17 - RANDOM NUMBER
    POSITIONS 18 THROUGH 19 - FRAME NUMBER
    POSITIONS 20 THROUGH 30 - SUB-FRAME NUMBER
    POSITIONS 31 THROUGH 41 - ITEM NUMBER WITHIN

THE SUB-FRAME
    EACH COLUMN OF NUMBERS IS RIGHT JUSTIFIED

PRESS ANY KEY TO CONTINUE
```

---

## RHC SAMPLE SELECTION

### Purpose

This program will randomly group and select for two and three stage samples 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 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 to be created. The universe in the data file may not exceed 6,000 units. In order to process such a large universe of primary (or secondary) units, the program will temporarily use disk space of the current directory. For each unit in the universe, the program will use 43 bytes of disk space. The temporary file is named TEMP2.RHC and is erased at the end of the program.

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 universe of hospitals for each state when performing the first phase (i.e. state sampling) of the RHC sample. 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. Once the sampled hospitals have been identified, the user must know 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 Queries

**WHEN PARTITIONING THE PRIMARY/SECONDARY UNITS, IF YOU WANT TO SEED THE GENERATOR, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY?**

The program allows a seed number to be entered by the user when generating the groups of units. 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 printed as part of the output. This option allows the user to obtain the same groups of primary units in subsequent computer runs if the same seed value is supplied at this query.

**WHEN SELECTING THE PRIMARY/SECONDARY UNIT SAMPLE, IF YOU WANT TO SEED THE GENERATOR, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY?**

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. 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 printed as part of the output. This option allows the user to obtain the same groups of primary units in subsequent computer runs if the same seed value is supplied at this query.

**IS THIS A SAMPLE OF <P>RIMARY OR <S>ECONDARY UNITS?**

This program may be used to generate a random sample in either a two or three stage sample. If it is a three stage sample, the user must use the results of the primary sample selection to generate a data file of secondary information only for the primary units selected and then run the program again with this new data file.

**ENTER NAME OF FILE CONTAINING INFORMATION ON ALL OF YOUR PRIMARY UNITS, ELSE ENTER QUIT?**

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. schools in Jefferson County), and a size value for the unit (e.g. number of students 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 students in a county may be the best size value for a sample of schools, the square footage of classroom space may be the best size indicator in another sample. The size value used impacts on the variance in the appraisal. However, the sample is not invalidated by using less than the best size indicator. The same size value category must be used throughout the data file. The user, for example, may not use the student population for one county and the classroom square footage for another county.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

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 sub-units 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.

**ENTER THE NUMBER OF PRIMARY UNITS TO BE SAMPLED?**

The user enters the sample size at the particular level (i.e. primary or secondary stage). 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.

**ENTER THE NAME OF THE OUTPUT FILE FOR THIS PROGRAM?**

The sample drawn from the universe of units will be output to a disk file using the name entered at this prompt. The primary file name may be up to eight characters and may have and may have an extension of up to three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be : "A:FILENAME.JUN".

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?**

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**ENTER NAME FOR FILE TO RECORD THE RESULTS?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name can be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.NBR" or "A:\RHC\FILE1.NBR" if it were to a directory on the "A" drive.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

## DO YOU WANT <A>LL OR ONLY <S>AMPLE PRIMARY UNITS PRINTED?

The user has the option of displaying the units that were randomly placed into each grouping and then the unit sampled from each group (i.e. the ALL selection) or displaying only the unit sampled from each group (i.e. the SAMPLE selection). The user only needs to enter **A** (all selection) or **S** (sample selection).

An example of the input screens is shown on page 2-48. The user entries are shown in **boldface**. The resultant output of the program is shown on pages 2-49 through 2-51.

### Program Output

If the ALL option was selected, then the following information will be displayed as part of the output.

<b>NAME OF INPUT FILE</b>	The file name entered by the user containing data for each unit in the universe.
<b>GROUP #</b>	There are groups formed randomly by the program. The total number of groups is equal to the number of units the user entered as the sample size.
<b>PRIMARY UNIT IDENTIFICATION</b>	The description assigned by the user to each unit in the universe. A maximum of 30 characters is allowed per unit description.
<b>UNIT SIZE</b>	The size factor assigned by the user to the unit. The factor could be square footage, number of employees, or number of cases at each unit. This value is used as the basis for PPS sampling within each group.
<b>SECONDARY UNIVERSE</b>	The universe of units within the primary units. The universe could be the number of claims or vouchers at a particular location. If the user does not know the universe of all units, then a one may be entered for this universe. The sample selection at this level is based on the PPS for the size factor. Once the units for sampling have been identified, then the universe within each unit must be known for the final sample selection.

For both the ALL and SAMPLE options, the following information will be displayed.

<b>FIRST SEED NUMBER</b>	This is the seed number either entered by the user or generated by the computer. This seed number was used to randomly place the units into groups.
<b>SECOND SEED NUMBER</b>	This is the seed number either entered by the user or generated by the computer. This seed number was used to select the unit within each group using the PPS method.
<b>NUMBER OF PRIMARY UNITS IN THE POPULATION</b>	This is the quantity of units that the program identified from the data file as being in the universe.
<b>NUMBER OF PRIMARY UNITS SAMPLED</b>	This is the quantity of units requested by the user to be sampled.
<b>PRIMARY UNIT ID</b>	This is the description entered by the user for the unit that was selected as a sample item.
<b>SECONDARY UNIVERSE</b>	The secondary universe entered in the data file for the unit being sampled.
<b>PRIMARY UNIT SIZE</b>	The size factor entered in the data file for the unit being sampled.
<b>GROUP SIZE</b>	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.
<b>UNITS IN GROUP</b>	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.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
GENERATION OF RHC SAMPLE  
VERSION: MONTH YEAR

DATE: 03/22/93 TIME: 18:35

WHEN PARTITIONING THE PRIMARY/SECONDARY UNITS, IF YOU WANT TO SEED THE GENERATOR, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY. **10**

WHEN SELECTING THE PRIMARY/SECONDARY UNIT SAMPLE, IF YOU WANT TO SEED THE GENERATOR, ENTER THE SEED NUMBER; ELSE HIT THE ENTER KEY. **20**

IS THIS A SAMPLE OF <P>RIMARY OR <S>ECONDARY UNITS (ENTER P OR S)? **P**

ENTER NAME OF FILE CONTAINING INFORMATION ON ALL OF YOUR PRIMARY UNITS, ELSE ENTER QUIT? **B:COUNTY.UNV**

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
GENERATION OF RHC SAMPLE  
VERSION: MONTH YEAR

DATE: 03/22/93 TIME: 18:35

ENTER THE NUMBER OF PRIMARY UNITS TO BE SAMPLED **8**

ENTER THE NAME OF THE OUTPUT FILE FOR THIS PROGRAM: **B:COUNTY.RND**

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

DO YOU WANT <A>LL OR ONLY <S>AMPLE PRIMARY UNITS PRINTED? **A**

DATE: 03/22/93

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 GENERATION OF RHC SAMPLE  
 VERSION: MONTH YEAR  
 NAME OF INPUT FILE: B:\COUNTY.UNV

TIME: 18:35

GROUPS OF PRIMARY UNITS

\*\*\*\*\* GROUP 1 \*\*\*\*\*

PRIMARY UNIT IDENTIFICATION =====	PRIMARY UNIT SIZE =====	SECONDARY UNIVERSE =====
COUNTY THIRTEEN	493	1
COUNTY FORTY	182	1
COUNTY TEN	166	1
COUNTY TWELVE	179	1
COUNTY THIRTY-ONE	135	1
GROUP TOTALS: 5	1,155	5

\*\*\*\*\* GROUP 2 \*\*\*\*\*

PRIMARY UNIT IDENTIFICATION =====	PRIMARY UNIT SIZE =====	SECONDARY UNIVERSE =====
COUNTY FOUR	127	1
COUNTY ELEVEN	252	1
COUNTY THIRTY-SIX	350	1
COUNTY THIRTY-TWO	662	1
COUNTY TWENTY-THREE	5,202	1
GROUP TOTALS: 5	6,593	5

\*\*\*\*\* GROUP 3 \*\*\*\*\*

PRIMARY UNIT IDENTIFICATION =====	PRIMARY UNIT SIZE =====	SECONDARY UNIVERSE =====
COUNTY THIRTY-THREE	344	1
COUNTY SEVEN	1,226	1
COUNTY THIRTY-SEVEN	600	1
COUNTY FOURTEEN	93	1
COUNTY FIFTEEN	209	1
GROUP TOTALS: 5	2,472	5

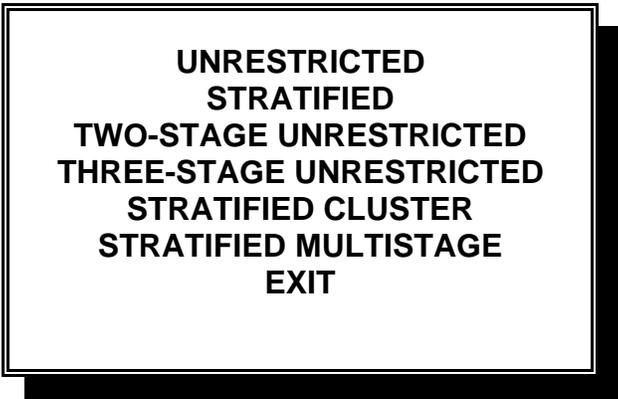
\*\*\*\*\* GROUP 4 \*\*\*\*\*

PRIMARY UNIT IDENTIFICATION =====	PRIMARY UNIT SIZE =====	SECONDARY UNIVERSE =====
COUNTY TWENTY-NINE	1,772	1
COUNTY TWENTY-EIGHT	455	1
COUNTY THIRTY-EIGHT	260	1
COUNTY THIRTY	250	1
COUNTY SEVENTEEN	1,871	1
GROUP TOTALS: 5	4,608	5

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 03/22/93	GENERATION OF RHC SAMPLE	
	VERSION: MONTH YEAR	
	TIME: 18:35	
	NAME OF INPUT FILE: B:\COUNTY.UNV	
GROUPS OF PRIMARY UNITS		
***** GROUP 5 *****		
PRIMARY UNIT IDENTIFICATION	PRIMARY UNIT SIZE	SECONDARY UNIVERSE
=====	=====	=====
COUNTY NINE	21,342	1
COUNTY THIRTY-FOUR	175	1
COUNTY THIRTY-NINE	502	1
COUNTY SIXTEEN	163	1
COUNTY EIGHTEEN	2,008	1
GROUP TOTALS: 5	24,190	5
***** GROUP 6 *****		
PRIMARY UNIT IDENTIFICATION	PRIMARY UNIT SIZE	SECONDARY UNIVERSE
=====	=====	=====
COUNTY THREE	2,086	1
COUNTY TWENTY-FOUR	2,557	1
COUNTY FIVE	682	1
COUNTY NINETEEN	253	1
COUNTY TWENTY-FIVE	1,770	1
GROUP TOTALS: 5	7,348	5
***** GROUP 7 *****		
PRIMARY UNIT IDENTIFICATION	PRIMARY UNIT SIZE	SECONDARY UNIVERSE
=====	=====	=====
COUNTY SIX	307	1
COUNTY TWENTY-TWO	2,869	1
COUNTY TWO	576	1
COUNTY TWENTY	300	1
COUNTY EIGHT	301	1
GROUP TOTALS: 5	4,353	5
***** GROUP 8 *****		
PRIMARY UNIT IDENTIFICATION	PRIMARY UNIT SIZE	SECONDARY UNIVERSE
=====	=====	=====
COUNTY ONE	2,990	1
COUNTY TWENTY-ONE	2,754	1
COUNTY THIRTY-FIVE	927	1
COUNTY TWENTY-SEVEN	651	1
COUNTY TWENTY-SIX	358	1
GROUP TOTALS: 5	7,680	5

DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES				
DATE: 03/22/93	GENERATION OF PRIMARY UNIT SAMPLE			TIME: 18:35
NAME OF OUTPUT FILE: COUNTY.RND				
FIRST SEED NUMBER: 10.0000		SECOND SEED NUMBER: 20.0000		
NUMBER OF PRIMARY UNITS IN THE POPULATION: 40				
NUMBER OF PRIMARY UNITS SAMPLED: 8				
PRIMARY UNIT ID	SECONDARY UNIVERSE	PRIMARY UNIT SIZE	GROUP SIZE	UNITS IN GROUP
=====	=====	=====	=====	=====
COUNTY TWELVE	1	179	1,155	5
COUNTY TWENTY-THREE	1	5,202	6,593	5
COUNTY SEVEN	1	1,226	2,472	5
COUNTY TWENTY-EIGHT	1	455	4,608	5
COUNTY NINE	1	21,342	24,190	5
COUNTY THREE	1	2,086	7,348	5
COUNTY EIGHT	1	301	4,353	5
COUNTY TWENTY-SIX	1	358	7,680	5

## OVERVIEW



**UNRESTRICTED  
STRATIFIED  
TWO-STAGE UNRESTRICTED  
THREE-STAGE UNRESTRICTED  
STRATIFIED CLUSTER  
STRATIFIED MULTISTAGE  
EXIT**

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

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

This package attempts to offer the user a variety of appraisal methodologies so as to allow the user a variety of options in designing and performing a statistical sample. A brief example for using each program is given below. A detailed explanation of how to use each module is described later in this section.

### UNRESTRICTED

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

## **STRATIFIED**

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

## **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 time cards mentioned above could be located throughout the United States. The reviewer could randomly select locations throughout the organization. Then at the selected locations, perform a sample of time cards. 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 time card 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 time cards.

## **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 time cards again, once the reviewer has arrived at the sampled location, it may take only a short amount of additional time to review all time cards 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 time cards.

## **STRATIFIED MULTISTAGE**

This methodology is similar to the "Stratified Cluster" with the exception that not all items in the sub-universe are reviewed. In the example discussed above, the assumption was made that the reviewer had sufficient time and resources to analyze all of the time 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.

These appraisal programs provide the correct statistical results only when the proper sample design has been executed. If an unrestricted sample of time cards 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 Queries

#### 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 may be entered with or without commas. For example, one million may be entered as 1000000 or 1,000,000.

#### SAMPLE SIZE?

The sample size is the quantity of items selected by the user from the universe. This number also may be entered with commas.

#### NUMBER OF SAMPLE ITEMS WITH THE DESIRED CHARACTERISTIC(S)?

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 ("desired characteristics"). Depending on the purpose of the appraisal, the user would enter the number of items meeting the criteria or the number of items failing to meet the criteria. For example, if the user was looking at 100 documents to see if the documents had the proper approval signature, then the desired characteristic 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 desired characteristic(s) as 88. The user could also enter 12 (100 - 88) if the purpose of the appraisal was to estimate what percentage of the universe of documents did not have the approval signature. The program will also evaluate samples that have:

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

### **OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN?**

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

### **ENTER OUTPUT FILE NAME FOR ATTRIBUTE APPRAISAL?**

If the user selects the disk for output, then the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILEONE.APR**" or "**A:\AUDIT\OUTFILE1.APR**" if it were going to a directory on the "A" drive.

By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the query screen appears on page 3-7. The user entries are in **boldface**. The resultant appraisal output, including the data entered by the user, appears on page 3-7.

## Program Output

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

<b>PROJECTED QUANTITY IN UNIVERSE</b>	The proportion of sample items with the desired characteristics multiplied by the universe size.
<b>PERCENT</b>	The proportion of sample items with the desired characteristics displayed as a percentage.
<b>CONFIDENCE LEVELS</b>	The confidence associated with the ability of the corresponding interval to contain the true proportion (or total number in universe).
<b>LOWER LIMIT</b>	The lower boundary of the confidence interval. The limit is shown as both a number and percentage of the universe.
<b>UPPER LIMIT</b>	The upper boundary of the confidence interval. The limit is shown as both a number and percentage of the universe.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 SINGLE-STAGE ATTRIBUTE APPRAISAL  
 VERSION: MONTH YEAR

DATE: 07/27/92 TIME: 08:47

NAME OF AUDIT/REVIEW? **TEST FOR ATTRIBUTE APPRAISAL**

UNIVERSE SIZE? **10,000**

SAMPLE SIZE? **120**

NUMBER OF SAMPLE ITEMS WITH DESIRED CHARACTERISTIC(S)? **12**

OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ?

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 SINGLE-STAGE ATTRIBUTE APPRAISAL  
 AUDIT: TEST FOR ATTRIBUTE APPRAISAL

DATE: 07/27/92 TIME: 08:47

UNIVERSE SIZE		10,000
SAMPLE SIZE		120
DESIRED CHARACTERISTIC(S)		
QUANTITY IDENTIFIED IN SAMPLE		12
PROJECTED QUANTITY IN UNIVERSE	1,000	
PERCENT		10.000%
CONFIDENCE LIMITS		
	90% CONFIDENCE LEVEL	95% CONFIDENCE LEVEL
LOWER LIMIT - QUANTITY	590	530
PERCENT	5.897%	5.303%
UPPER LIMIT - QUANTITY	1,566	1,678
PERCENT	15.664%	16.776%

---

---

## 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. The program will accept a maximum of 20 strata for an appraisal.

### Input Queries

#### 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.

#### IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL, ENTER NAME OF FILE, ELSE HIT THE <ENTER> KEY?

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 desired characteristic(s). 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, simply hitting the Enter key will cause the program to assume that the input will be entered from the keyboard.

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 universe of items from which a sample was drawn in a stratum.

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

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

### **NUMBER OF STRATA IN THIS APPRAISAL?**

If the user did not enter a file name at the previous prompt, then the user will be queried as to the number of strata to be appraised.

```

                                CHARACTER-
STRAT UNIVERSE SAMPLE ISTIC(S)
=====

```

If the values are being entered from the keyboard, then the above header will be displayed on the screen. The user should first enter the value for the universe size for stratum 1. The value may be entered with or without commas. After entering the universe size, the user must hit the Enter key and the cursor will move to the "Sample" column. The user then enters the sample size (commas are optional) and hits the Enter key again before entering the number of items with the desired characteristic(s). The process will continue until the values for all the strata have been entered.

After the values have been entered for a stratum, the program will test to ensure that the sample size is not greater than the universe size and the number of items meeting the desired characteristic(s) is not greater than the sample size. If either condition exists, the user will be prompted to re-enter the three pieces of data for the stratum. Additional editing may be made after all values for the appraisal have been entered.

### **ENTER STRATUM NUMBER TO BE EDITED(ELSE HIT <ENTER>)?**

If any of the values previously entered need to be edited then the user may use this routine to identify the stratum that needs to be changed. This query will continue until the user hits the Enter key without entering a value.

---

**ENTER UNIVERSE SIZE FOR STRATUM?**

If a number was entered to the prior query, then the program prompts for the universe value for the stratum being edited. If the universe size does not need to be changed, then the user may merely hit the Enter key and the current universe size will be retained. If a new value is entered, it may be entered with or without commas.

**ENTER SAMPLE SIZE FOR STRATUM?**

If the user had indicated a stratum to be edited, then at this prompt the user would enter the new sample size value. If the sample size does not need to be changed, then the user simply hits the Enter key and the current value will be retained. Commas may be used in entering the sample size.

**ENTER NUMBER OF ITEMS WITH DESIRED CHARACTERISTICS?**

If the user had indicated a stratum to be edited, then at this prompt the user would enter the number of sample items reviewed that met the evaluation criteria. If the value does not need to be edited, then the user simply hits the Enter key and the current value will be retained. Commas may be used in entering the value.

**IF YOU WOULD CARE TO SAVE THE SET OF VALUES, ENTER THE FILE NAME (ELSE HIT <ENTER>)?**

If any of the values were entered from the keyboard, then the program will prompt the user for a file name for saving the values. If no name is entered, then the values will not be saved after the appraisal is made.

**OUTPUT APPRAISAL TO <D>ISK,<P>RINTER OR <S>CREEN?**

The program allows for three output options. The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**ENTER OUTPUT FILE NAME FOR ATTRIBUTE APPRAISAL?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.APR" or "A:\AUDIT\OUTFILE1.APR" if it were going to a directory on the "A" drive.

By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the query screen appears on page 3-12. The user entries are in **boldface**. The output from the resultant appraisal appears on page 3-12.

### **Program Output**

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

<b>RATIO</b>	The percentage of sample items in each stratum with the desired characteristics.
<b>PROJ. ITEMS IN UNIVERSE</b>	The result of applying the stratum ratio to the stratum universe. This is also calculated for the total universe.
<b>PRECISION</b>	The confidence interval half-width expressed as a percentage. Precision is calculated for each stratum and the universe as a whole.
<b>LOWER LIMIT</b>	The lower boundary of the confidence interval. The limit is shown as both a number and a percentage of the universe.
<b>UPPER LIMIT</b>	The upper boundary of the confidence interval. The limit is shown as both a number and a percentage of the universe.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
STRATIFIED ATTRIBUTE APPRAISAL  
VERSION: MONTH YEAR

DATE: 07/24/92 TIME: 07:26

NAME OF AUDIT/REVIEW? **TEST FOR STRATIFIED APPRAISAL**

IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL, ENTER NAME OF FILE  
ELSE HIT THE <ENTER> KEY?

NUMBER OF STRATA IN THIS APPRAISAL? **2**

STRAT	UNIVERSE	SAMPLE	CHARACTER- ISTIC (S)	STRAT	UNIVERSE	SAMPLE	CHARACTER- ISTIC (S)
=====	=====	=====	=====	=====	=====	=====	=====
1	1,000	100	2	2	1,000	100	4

ENTER STRATUM NUMBER TO BE EDITED (ELSE HIT <ENTER>)?

IF YOU WOULD CARE TO SAVE THIS SET OF VALUES, ENTER A FILE NAME  
(ELSE HIT <ENTER>)?

OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
STRATIFIED ATTRIBUTE APPRAISAL

DATE: 07/24/92 TIME: 07:26

AUDIT/REVIEW: TEST FOR STRATIFIED APPRAISAL

STRAT	SAMPLE	*ITEMS**	**RATIO*	*UNIVERSE*	PROJ. ITEMS IN UNIVERSE	PRECISION AT 90% CL	PRECISION AT 95% CL
=====	=====	=====	=====	=====	=====	=====	=====
1	100	2	2.000%	1,000	20	2.196%	2.616%
2	100	4	4.000%	1,000	40	3.073%	3.662%
TOTALS							
2	200	6	3.000%	2,000	60	1.888%	2.250%
LOWER LIMIT - QUANTITY						22	15
PERCENT						1.112%	0.750%
UPPER LIMIT - QUANTITY						98	105
PERCENT						4.888%	5.250%

---

---

## 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 20 primary sampling units for an appraisal.

### Input Queries

#### 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.

#### IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL, ENTER NAME OF FILE, ELSE HIT THE <ENTER> KEY?

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 desired characteristic(s). 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, simply hitting the Enter key will cause the program to assume that the input will be entered from the keyboard.

If a file name is entered, the values obtained from the file will be displayed on the screen for review by the user.

**1000 100 2**

Explanation:

**1000 -** This is the number of items in the primary unit from which a sample was drawn.

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

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

### NUMBER OF PRIMARY UNITS SAMPLED IN THIS APPRAISAL?

If the user did not enter a file name at the previous prompt, then the user will be queried as to the number of primary units that were sampled for appraisal.

**CHARACTER-**  
**PRIME UNIVERSE SAMPLE IISTIC(S)**  
 =====

If the values are being entered from the keyboard, then the above header will be displayed on the screen. The user should first enter the value for the universe size for primary unit 1. The value may be entered with or without commas. After entering the universe size, the user must hit the Enter key and the cursor will move to the "Sample" column. The user then enters the sample size (commas are optional) and hits the <Enter> key again before entering the number of items with the desired characteristic(s). The process will continue until the values for all the primary units sampled have been entered.

After the values have been entered for a primary unit, the program will test to ensure that the sample size is not greater than the universe size and the number of items meeting the desired characteristic(s) is not greater than the sample size. If either condition exists, the user will be prompted to re-enter the three pieces of data for the primary unit. Additional editing may be made after all values for the appraisal have been entered.

**ENTER PRIMARY UNIT TO BE EDITED (ELSE HIT <ENTER>)?**

If any of the values previously entered need to be edited then the user may use this routine to identify the primary unit that needs to be changed. This query will continue until the user hits the Enter key without entering a value.

**ENTER THE UNIVERSE OF SECONDARY UNITS?**

If a number was entered to the prior query, then the program prompts for the universe size for the primary unit being edited. If the universe size does not need to be changed, then the user may merely hit the Enter key and the current universe value will be retained. If a new value is entered, it may be entered with or without commas.

**ENTER SAMPLE SIZE FOR THIS PRIMARY UNIT?**

If the user had indicated a primary unit to be edited, then at this prompt the user would enter the new sample size value. If the sample size does not need to be changed, then the user simply hits the Enter key and the current value will be retained. Commas may be used in entering the sample size.

**ENTER NUMBER OF ITEMS WITH DESIRED CHARACTERISTICS?**

If the user had indicated a primary unit to be edited, then at this prompt the user would enter the number of sample items reviewed that met the evaluation criteria. If the value does not need to be edited, then the user simply hits the Enter key and the current value will be retained. Commas may be used in entering the value.

**IF YOU WOULD CARE TO SAVE THE SET OF VALUES, ENTER THE FILE NAME (ELSE HIT <ENTER>)?**

If any of the values were entered from the keyboard, then the program will prompt the user for a file name for saving the values. If no name is entered, then the values will not be saved after the appraisal is made.

**ENTER THE UNIVERSE OF PRIMARY UNITS?**

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 in a state. From a universe of 58 counties, one might sample 10 counties. The response to this query would be 58.

### **ENTER UNIVERSE OF SECONDARY UNITS. IF UNKNOWN ENTER 0 (ZERO)?**

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.

### **OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN?**

The program allows for three output options. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

### **ENTER THE NAME OF THE FILE TO RECORD THE RESULTS?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.APR" or "A:\AUDIT\OUTFILE1.APR" if it were going to a directory on the "A" drive.

By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number

between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the query screen appears on page 3-18. The user entries are in **boldface**. An example of the output appears on page 3-19.

## Program Output

The program will print the data supplied by the user (universe size, sample size and number of items with desired characteristics) for each primary unit. In addition, the following pieces of information will be displayed for each primary unit:

**RATIO**            The percentage of sample items in each primary unit with the desired characteristics.

**PROJECTED**    If the total number of secondary items in the universe was entered, then the projected number of items meeting the desired characteristic(s) for each primary unit will be displayed.

The following overall statistics will also be displayed.

**OVERALL RATIO**    This is the estimate of the universe proportion, stated as a percentage.

**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**        A measurement of the closeness of the sample estimate and the corresponding unknown universe value. The precision is calculated by multiplying the standard error by the appropriate factor ("z" value) corresponding to the desired confidence level. The results are shown at both the 90 and 95 percent confidence levels in terms of a percentage. If the user entered a value for the universe of secondary items (other than zero), the output will also display the precision as a quantity of items. This quantity is obtained by multiplying the secondary universe by the precision percentage.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 TWO STAGE UNRESTRICTED APPRAISAL  
 VERSION: MONTH YEAR

DATE: 08/21/92

TIME: 16:34

NAME OF AUDIT/REVIEW? **TEST FOR TWO STAGE ATTRIBUTE APPRAISAL**IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL, ENTER NAME OF FILE  
ELSE HIT THE <ENTER> KEY?NUMBER OF PRIMARY UNITS SAMPLED IN THIS APPRAISAL? **10**

PRIME	UNIVERSE	SAMPLE	CHARAC- TERISTIC(S)	PRIME	UNIVERSE	SAMPLE	CHARAC- TERISTIC (S)
=====	=====	=====	=====	=====	=====	=====	=====
1	50	10	4	2	65	13	5
3	45	9	2	4	48	10	3
5	52	10	5	6	58	12	3
7	42	8	3	8	66	13	4
9	40	8	2	10	56	11	4

ENTER PRIMARY UNIT TO BE EDITED (ELSE HIT &lt;ENTER&gt;)?

IF YOU WOULD CARE TO SAVE THIS SET OF VALUES, ENTER A FILE NAME  
(ELSE HIT <ENTER>)?ENTER THE UNIVERSE OF PRIMARY UNITS? **90**ENTER THE UNIVERSE OF SECONDARY UNITS. IF UNKNOWN ENTER 0 (ZERO)? **4,500**OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

DEPARTMENT OF HEALTH & HUMAN SERVICES						
OIG - OFFICE OF AUDIT SERVICES						
TWO STAGE ATTRIBUTE APPRAISAL						
DATE: 08/21/92						TIME: 16:34
AUDIT: TEST FOR MULTISTAGE ATTRIBUTE APPRAISAL						
PRIMARY UNIT	UNIVERSE	SAMPLE SIZE	SAMPLE ITEMS WITH CHARACTERISTIC (S)	RATIO	PROJECTED	
=====	=====	=====	=====	=====	=====	
1	50	10	4	40.00%	20	
2	65	13	5	38.46%	25	
3	45	9	2	22.22%	10	
4	48	10	3	30.00%	14	
5	52	10	5	50.00%	26	
6	58	12	3	25.00%	15	
7	42	8	3	37.50%	16	
8	66	13	4	30.77%	20	
9	40	8	2	25.00%	10	
10	56	11	4	36.36%	20	
TOTALS	522	104	35			
OVERALL TOTALS:						
90	4,500			35.26%	1,587	
STANDARD ERROR				+/- 3.67%		
PRECISION						
AT 90% CONFIDENCE LEVEL				+/- 6.04%	272	
AT 95% CONFIDENCE LEVEL				+/- 7.20%	324	

---

## 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 attribute. The program will accept a maximum of 50 first stage units and 50 second stage units within each first stage unit (i.e.  $50 \times 50 = 2500$  second stage units).

### Input Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

Prior to executing this program, the user must create a data file that contains the results of the three stage attribute sample. If the user has not created such a file, the word **QUIT** is entered to exit the program. 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 DOS (ASCII) 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. WordPerfect) or storing the data in a spreadsheet software application (e.g. Quattro Pro) as a print file. The data file needs to be formatted in the following manner:

```
NORTHWEST REGION 50 6  
LOCATION-1      47 9 3
```

Explanation:

**NORTHWEST REGION** - 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.

**50** - is the universe of secondary units within the primary unit.

**6** - is the number of secondary units that are being sampled in the primary unit. If the number of sample units is greater than the universe, the program will assume the user entered the universe and sample in reverse order and swap the data.

- LOCATION-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 attribute of interest to the user.

The last piece of data (e.g. 6) 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. 6) to be lines with secondary unit information. Each line of data must contain information on only one primary unit or one secondary unit.

#### **ENTER THE NUMBER OF PRIMARY UNITS IN THE UNIVERSE?**

In the preceding illustration, this would be the number of regions in the universe. This value must be known.

#### **ENTER THE NUMBER OF UNIVERSE THIRD STAGE UNITS. IF UNKNOWN ENTER 0 (ZERO).**

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.

#### **DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer), or **S** (screen).

### ENTER THE NAME OF THE FILE TO RECORD THE RESULTS?

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILENAME.EXT**" or "**A:\AUDIT\OUTFILE1.APR**" if it were going to a directory on the "A" drive.

By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILENAME.EXT** to have the results displayed on the screen or **TYPE A:FILENAME.EXT > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

If the user selects the printer for output, then this query will appear. Users that have computers with the ability to access multiple parallel ports or connections through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter a port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Example of the query screens appears on page 3-25. The user entries are in **boldface**. The related output is shown on pages 3-26 and 3-27.

## 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 attribute of interest. Totals are printed 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 attribute 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 following information is also presented:

**OVERALL POINT ESTIMATE OF THE PROPORTION**

This is the estimate of the universe proportion having the attribute 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**

This is the estimate of the total number of elements in the universe having the attribute 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.

**LOWER LIMIT FOR PROPORTION**

The lower limit of the 90% and 95% confidence interval for the universe proportion. The lower limit is based on a normal approximation and is equal to the point estimate of proportion less the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).

**UPPER LIMIT FOR PROPORTION**

The upper limit of the 90% and 95% confidence interval for the universe proportion. This limit is also based on a normal approximation and is equal to the point estimate of proportion plus the "z" value multiplied by the standard error.

**LOWER LIMIT FOR TOTAL**

The lower limit of the 90% and 95% confidence interval for the universe total number. The lower limit is based on a normal approximation and is equal to the point estimate of the universe total less the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).

**UPPER LIMIT FOR TOTAL**

The upper limit of the 90% and 95% confidence interval for the universe total number. This limit is also based on a normal approximation and is equal to the point estimate of universe total plus the "z" value multiplied by the standard error.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 THREE STAGE ATTRIBUTE APPRAISAL  
 VERSION: MONTH YEAR

DATE: 05/28/92 TIME: 12:30

ENTER THE NAME OF YOUR DATA FILE OR QUIT? **A:3STAGE.DAT**

ENTER THE NUMBER OF PRIMARY UNITS IN THE UNIVERSE? **12**

ENTER THE NUMBER OF UNIVERSE THIRD STAGE UNITS.  
 IF KNOWN ENTER 0 (ZERO)? **48100**

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)? **1**

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 THREE STAGE ATTRIBUTE APPRAISAL  
 NAME OF DATA FILE: A:3STAGE.DAT

DATE: 05/28/92 TIME: 12:30

FIRST STAGE SECOND STAGE =====	NEXT STAGE UNIVERSE =====	SAMPLE SIZE =====	MEETING CRITERIA =====	RATIO =====
REGION 1	50	6		
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	10	33.33%
HOSPITAL 6	50	10	1	10.00%
TOTALS	285	56	14	

DEPARTMENT OF HEALTH & HUMAN SERVICES					
OIG - OFFICE OF AUDIT SERVICES					
THREE STAGE ATTRIBUTE APPRAISAL					
DATE: 05/28/92					TIME: 12:30
NAME OF DATA FILE: A:3STAGE.DAT					
FIRST STAGE SECOND STAGE	NEXT STAGE UNIVERSE	SAMPLE SIZE	MEETING CRITERIA	RATIO	
=====	=====	=====	=====	=====	
REGION 1	90	6			
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%	
TOTALS	285	56	14		
REGION 2	110	6			
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%	
TOTALS	363	73	18		
REGION 3	85	6			
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%	
TOTALS	269	55	15		
REGION 4	120	6			
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%	
TOTALS	399	80	20		

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 05/28/92	THREE STAGE ATTRIBUTE APPRAISAL	TIME: 12:30
NAME OF INPUT FILE: A:3STAGE.DAT		
TOTAL NUMBER OF THIRD STAGE UNITS IN UNIVERSE	48,100	
OVERALL TOTALS	UNIVERSE	SAMPLED
=====	=====	=====
FIRST STAGE	12	4
SECOND STAGE	405{}	24
THIRD STAGE	1,316{}	264
SAMPLED ITEMS MEETING CRITERIA		67
{ } UNIVERSE SIZES FOR THE SECOND AND THIRD STAGES REPRESENT THE UNIVERSES FROM THE SAMPLED PRIOR STAGE.		
OVERALL POINT ESTIMATE OF THE PROPORTION		35.81%
OVERALL STANDARD ERROR (PROPORTION)		5.46%
OVERALL POINT ESTIMATE OF THE UNIVERSE TOTAL		17,224
OVERALL STANDARD ERROR (TOTAL)		2,625
CONFIDENCE LEVEL	90 PERCENT	95 PERCENT
LOWER LIMIT FOR PROPORTION	26.83%	25.11%
UPPER LIMIT FOR PROPORTION	44.79%	46.51%
LOWER LIMIT FOR TOTAL	12,906	12,079
UPPER LIMIT FOR TOTAL	21,543	22,370

---

---

## 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 maximum of 20 strata and a sample of no more than 100 primary units within each stratum (i.e.  $20 \times 100 = 2000$  primary units). This procedure is used when all secondary items in a sampled primary unit are reviewed for sampling purposes.

### Input Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

Prior to executing this program, the user must create a data file that contains the results of the stratified cluster attribute sample. If the user has not created such a file, the word **QUIT** is entered to exit the program. 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 DOS (ASCII) 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. WordPerfect) or storing the data in a spreadsheet software application (e.g. Lotus 1-2-3) as a print file. The data file needs to be formatted in the following manner:

```
STATE UNIVERSITIES 415 2,500
STATE SCHOOL 1      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 universe of primary units within the stratum.

**2,500** - is the universe of secondary units in the stratum.

- STATE SCHOOL 1** - 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 attribute of interest to the user.

At the end of all data lines for each stratum, the user must enter a line with the entry **3E33**. This is a sentinel that indicates to the program that it has reached the end of a stratum.

**DO YOU WANT OUTPUT TO PROVIDE <C>OMplete OR  
<S>UMMARY INFORMATION?**

The user has the option of displaying the details of the data file in columnar form with appropriate headings as part of the output (complete option) or only the stratum totals (summary option).

**DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR  
<S>CREEN?**

The user may select the output to be sent to disk, printer or screen. The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer), or **S** (screen).

**NAME OF DISK FILE FOR APPRAISAL OUTPUT?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILENAME.EXT**" or '**A:\AUDIT\OUTFILE1.APR**' if it were going to a directory on the "A" drive.

By sending the output to disk, the user may transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILENAME.EXT** to have the results displayed on the screen or **TYPE A:FILENAME.EXT > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Users that have computers with the ability to access multiple parallel ports or connections through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter a port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Examples of the query screen appear on page 3-33. The user entries are in **boldface**. The related output is also shown on page 3-33.

### **Program Output**

The first section of the program output contains the name of the data file along with the contents of this file in at least summary form. If the complete option was selected, then for each cluster in the file, the output displays the percentage of secondary items having the attribute of interest. In addition, if the complete option was selected, 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 attribute of interest. The following information is also presented:

**OVERALL - PERCENT**      This is the estimate of the universe proportion having the attribute of interest.

**OVERALL - PROJECTED QUANTITY**      This is the estimate, expressed in terms of secondary units, of items having the attribute 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.

---

<b>OVERALL - STANDARD ERROR (#)</b>	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.
<b>LOWER - LIMIT (QUANTITY)</b>	The lower limit of the 90% and 95% confidence interval for the universe total number. The lower limit is based on a normal approximation and is equal to the point estimate of universe total less the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).
<b>UPPER - LIMIT (QUANTITY)</b>	The upper limit of the 90% and 95% confidence interval for the universe total number. The lower limit is based on a normal approximation and is equal to the point estimate of universe total plus the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).
<b>LOWER - LIMIT (PERCENT)</b>	The lower limit of the 90% and 95% confidence interval for the universe proportion. The lower limit is based on a normal approximation and is equal to the point estimate of universe percent less the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).
<b>UPPER - LIMIT (PERCENT)</b>	The upper limit of the 90% and 95% confidence interval for the universe proportion. The lower limit is based on a normal approximation and is equal to the point estimate of universe percent plus the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).
<b>PRECISION - (QUANTITY)</b>	A measurement of the closeness of the sample estimate of the universe total and the corresponding unknown universe value. For a 90% confidence interval, the user would be 90% confident that the estimated total number of items having the attribute of interest (OVERALL PROJECTED QUANTITY) would be within this number of the actual value. The precision quantity and percent are based on the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).

---

**PRECISION - (PERCENT)** - A measurement of the closeness of the sample estimate of the universe proportion and the corresponding unknown universe percent. For a 90% confidence interval, the user would be 90% confident that the estimated proportion of the attribute of interest (OVERALL PERCENT) would be within this percent of the actual proportion. The precision percent is based on the "z" value multiplied by the standard error, where the "z" value is equal to 1.64485 (90%) or 1.95996 (95%).

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 05/28/92          STRATIFIED CLUSTER ATTRIBUTE APPRAISAL          TIME: 12:30
                        VERSION: MONTH YEAR

ENTER NAME OF YOUR DATA FILE OR QUIT? A:\CLUSTER.DAT

DO YOU WANT THE OUTPUT TO PROVIDE <C>OMPLETE OR <S>UMMARY INFORMATION? S

DO YOU WANT THE RESULTS PRINT TO <D>ISK, <P>RINTER OR <S>CREEN? P

ENTER THE PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)? S
    
```

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 05/28/92          STRATIFIED CLUSTER ATTRIBUTE APPRAISAL          TIME: 12:30
                        NAME OF DATA FILE: A:\CLUSTER.DAT
    
```

STRATUM IDENTIFICATION	SAMPLE UNIVERSE	SAMPLE SIZE	MEETING CRITERIA	PROJECTED PERCENT	PROJECTED QUANTITY
STATE UNIVERSITIES	415	25			
PRIVATE UNIVERSITIES	168	10			
STRATA TOTALS	583	35			
CLUSTER UNIT TOTALS	3,500	200	57		
OVERALL TOTALS				29.05%	1,017
OVERALL STANDARD ERROR				1.32%	46
			90% CONFIDENCE LEVEL	95% CONFIDENCE LEVEL	
LOWER LIMIT - QUANTITY			941	926	
PERCENT			26.88%	26.47%	
UPPER LIMIT - QUANTITY			1,093	1,107	
PERCENT			31.22%	31.64%	
PRECISION - QUANTITY			76	90	
PERCENT			2.17%	2.59%	

---

## 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 Queries

**IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL,  
ENTER NAME OF FILE, ELSE HIT THE <ENTER> KEY?**

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 DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

For each stratum the user needs to use the following format.

**26.20 7.50 10,000**

Explanation:

**26.20** - This is the overall ratio for one stratum. The estimate is shown as a percentage.

**7.50** - This is the standard error for one stratum. The standard error is shown as a percentage.

**10,000** - 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.

The results for each stratum should be on a separate line and the values separated by one or more spaces.

### **NUMBER OF STRATA FOR THIS APPRAISAL?**

If the user selects to enter the data interactively, the program will prompt for the number of strata that will be appraised. The number entered must be between 2 and 20.

Once the number of strata is entered, column headings will appear and the user will be prompted to enter the overall ratio, standard error and the universe for each stratum.

### **ENTER STRATUM TO BE EDITED (ELSE HIT <ENTER>)?**

After the data for all strata have been entered, the program will allow the user to edit any of the values. At this prompt, the user will enter the appropriate number (between 1 and the total number of strata) to indicate the stratum that is to be edited. If the user only hits the Enter key, then the program assumes that the user has completed editing the data.

### **ENTER THE OVERALL RATIO?**

For the stratum identified in the prior query, the user may now enter a new overall ratio. If the user does not want to change the ratio, then the user only needs to hit the Enter key.

### **ENTER THE STANDARD ERROR?**

For the stratum identified in the earlier query, the user may now enter a new standard error. If the user does not want to change the standard error, then the user only needs to hit the Enter key.

### **ENTER THE UNIVERSE OF DETAIL ITEMS?**

For the stratum identified in the earlier query, the user may now enter a new universe of sample units. If the user does not want to change the universe, then the user only needs to hit the Enter key.

---

**IF YOU WANT TO SAVE THE SET OF VALUES, ENTER A FILE NAME (ELSE HIT <ENTER>?)**

If the data was entered interactively or any changes were made, then the user will be given the opportunity to save the information. The primary file name may be up to eight characters. The user may also add an extension to the file with a period and up to three characters after the period. If the user wishes to place the file in a directory other than the current directory, then a complete path must be given to that directory. For example, if the data is to be stored on the "A" drive, the user could enter "A:STRDATA.DAT".

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D**(disk), **P**(printer) or **S**(screen).

**ENTER NAME FOR FILE TO RECORD THE RESULTS**

If the user selects the disk for the output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.OUT" or "A:\AUDIT\OUTFILE1.APR" if it were going to a directory on the "A" drive.

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE FILEONE.OUT** to have the results displayed on the screen or **TYPE FILEONE.OUT > PRN** to have the results sent to the printer. The user could also retrieve the file through a word processing package such as WordPerfect.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)**

If the user selects the printer for the output, the program will prompt for a parallel port number. If the output is to be sent to LPT1, the user has the option of entering a 1 or just hitting the Enter key. The program tests to see if the number is in the range 1 through 4. However, the program will not test to see if the port number entered is valid for the particular workstation.

An example of the query screen is on page 3-38. The user entries are in **boldface**. The resultant output of the appraisal is shown on page 3-38.

## 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.

**ESTIMATED ATTRIBUTE** - This is a single estimate for all strata of the universe value of the characteristic(s) being measured. The estimate is stated as a percentage.

**STANDARD ERROR** - This is a measurement of the variation of the estimated attribute with respect to all possible estimates for this universe and these sample sizes.

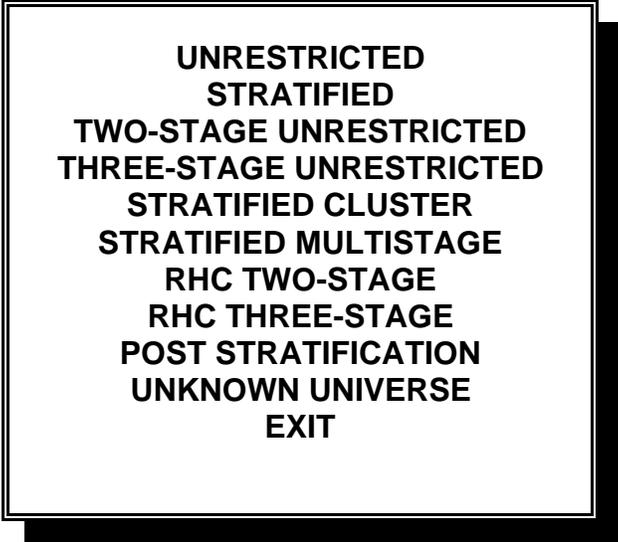
**CONFIDENCE LEVEL** - This indicates the confidence that the user has that the actual value of the measured characteristic will fall within the range from the lower to upper limits (confidence interval).

**LOWER LIMIT** - This is the lower bound of the confidence interval. It is calculated by subtracting the precision from the estimated attribute. The precision is based on the "z" value for the appropriate confidence level multiplied by the standard error. The "z" values are 1.64485 at the 90% confidence level and 1.95996 at the 95% confidence level.

**UPPER LIMIT** - This is the upper bound of the confidence interval. It is calculated by adding the precision to the estimated attribute. The precision is based on the "z" value for the appropriate confidence level multiplied by the standard error. The "z" values are 1.64485 at the 90% confidence level and 1.95996 at the 95% confidence level.



---

**OVERVIEW**

**UNRESTRICTED  
STRATIFIED  
TWO-STAGE UNRESTRICTED  
THREE-STAGE UNRESTRICTED  
STRATIFIED CLUSTER  
STRATIFIED MULTISTAGE  
RHC TWO-STAGE  
RHC THREE-STAGE  
POST STRATIFICATION  
UNKNOWN UNIVERSE  
EXIT**

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

This package attempts to offer the user a several appraisal methodologies so as to allow the user a variety of options in designing and performing a 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 a simple 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 impact 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.

## **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 sub-universe 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.

## **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. 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.

## **POST STRATIFICATION**

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**

The reviewer may have a situation where the size of the universe of items to be sampled is not known and can not 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.

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.

---

---

## 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 Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

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

**7483 289.99 43.00**

Explanation:

- 7483 -** This is a number assigned by the user. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.
- 289.99 -** This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.
- 43.00 -** If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of \$289.99 claimed by a vendor, only \$246.99 was actually owed, then the difference amount entered would be \$43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

The user has the option at the end of the file of adding the following line, "9999 3E33". This line serves as a sentinel. The line number may be any number selected by the user. In stratified samples, the program will use this line to indicate where data from one stratum ends. This line is optional in an unrestricted variable appraisal data file.

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data

file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

**THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

**ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR DATAFILE?**

The user selects the format that was used in creating the data file of values. Instead of "DATAFILE", the actual name of the data file entered in the previous response by the user will be displayed.

**DO YOU WANT <C>OMPLETE, <P>ARTIAL OR <S>UMMARY APPRAISAL?**

Only the first letter of the appraisal options needs to be entered. This query is usually for stratified samples, where the user may want to reduce printed output by having only a portion of the appraisal printed. In unrestricted random samples, the user may decide initially to appraise examined, audited and difference amounts. Subsequent appraisals of the sample after further review and modifications may be only for the difference amounts. In the latter case, the **Partial** selection would allow for such an appraisal.

**DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**NAME OF DISK FILE FOR APPRAISAL OUTPUT?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be entered as "A:FILEONE.APR."

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

While reading the data file, the program will display the name of the data file and the current record being processed. In displaying the current record, the program will limit the display to 50 characters.

After the program has read all of the values in the data file, a summary of the file will appear on the screen. If the user has selected the printer for output, the summary will also appear on the printer. At this point the user should reconcile the values to determine that the data file is complete and accurate.

### **CONTINUE PROCESSING WITH THE CURRENT FILE (YES/NO)?**

By entering the first letter of the appropriate response, the user may continue with the appraisal or skip the appraisal portion of the program.

### **ENTER THE UNIVERSE SIZE?**

The number entered represents the size of the universe from which the sample was drawn. The value may be entered with or without commas. This number will be used in estimating universe characteristics.

An example of a query screen is on page 4-11. The user entries are in **boldface**. The resultant output of the appraisal, including a summary of the data file, is shown on page 4-12.

## Program Output

For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed:

- MEAN** - The average value for the sample items appraised. It is obtained by summing the items in the sample and dividing the result by the number of items in the sample.
- UNIVERSE** - This is the quantity of the items from which the sample was drawn. The results of the sample will be projected to the universe using this value.
- STANDARD DEVIATION** - A measurement of the variation of the sample items about the average value (mean).
- STANDARD ERROR** - A measurement of the variation of the sample mean with respect to all possible means for this universe and sample size.
- SKEWNESS** - A measure of the symmetry of the frequency distribution of the sample items. Accounting universes are usually right skewed (majority of items have a low value while a few items have a high value).
- KURTOSIS** - A measure of the peakedness or flatness of the frequency distribution of the sample values.
- POINT ESTIMATE** - A single estimate for the universe total based on the sample mean multiplied by the universe size.
- CONFIDENCE LEVEL** - The confidence 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 -** This is the result of dividing the precision amount by the  
**PERCENT** point estimate and stating the result as a percentage.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 07/24/92          VARIABLES APPRAISAL          TIME: 10:40
                        VERSION: MONTH YEAR

ENTER THE NAME OF YOUR DATA FILE OR QUIT? B:\UNRESTR.DAT

                        THE ALLOWABLE FILE FORMATS ARE:
1 - EXAMINED VALUES ONLY      4 - EXAMINED AND AUDITED VALUES
2 - AUDITED VALUES ONLY      5 - EXAMINED AND DIFFERENCE VALUES
3 - DIFFERENCE VALUES ONLY    6 - AUDITED AND DIFFERENCE VALUES
                                0 - EXIT THE PROGRAM

ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR B:\AUDIT\OUTFILE1.NBR? 5

DO YOU WANT <C>COMPLETE, <P>ARTIAL OR <S>UMMARY APPRAISAL? C

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ?

=====NEW SCREEN=====

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 07/24/92          VARIABLES APPRAISAL          TIME: 10:40

SAMPLE      EXAMINED      NONZERO      TOTAL OF      TOTAL OF
SIZE        VALUE           DIFFS        DIFF VALUES  ADJ VALUES
1,246      457,360.00     1,246       456,422.00   938.00

CONTINUE PROCESSING WITH THIS CURRENT FILE (Y/N) ? Y

ENTER YOUR UNIVERSE SIZE? 10,000

```

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
VARIABLES APPRAISAL		
DATE: 07/24/92	DATA FILE USED: B:UNRESTR.DAT	
		TIME: 10:40
----- E X A M I N E D -----		
MEAN / UNIVERSE	367.06	10,000
STANDARD DEVIATION	427.60	
STANDARD ERROR	11.33	
SKEWNESS	1.09	
KURTOSIS	3.11	
POINT ESTIMATE	3,670,626	
	----90 PERCENT----	----95 PERCENT----
LOWER LIMIT	3,484,062	3,448,270
UPPER LIMIT	3,857,190	3,892,982
PRECISION AMOUNT	186,564	222,356
PRECISION PERCENT	5.08%	6.06%
----- A D J U S T E D -----		
MEAN / UNIVERSE	0.75	10,000
STANDARD DEVIATION	658.05	
STANDARD ERROR	17.44	
SKEWNESS	0.54	
KURTOSIS	2.82	
POINT ESTIMATE	7,528	
	----90 PERCENT----	----95 PERCENT----
LOWER LIMIT	-279,584	-334,666
UPPER LIMIT	294,641	349,722
PRECISION AMOUNT	287,113	342,194
PRECISION PERCENT	3,813.88%	4,545.56%
----- D I F F E R E N C E -----		
MEAN / UNIVERSE	366.31	10,000
STANDARD DEVIATION	426.91	
STANDARD ERROR	11.32	
SKEWNESS	1.09	
KURTOSIS	3.12	
POINT ESTIMATE	3,663,098	
	----90 PERCENT----	----95 PERCENT----
LOWER LIMIT	3,476,834	3,441,100
UPPER LIMIT	3,849,362	3,885,095
PRECISION AMOUNT	186,264	221,998
PRECISION PERCENT	5.08%	6.06%

---

## 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. If no variation exists, then there is no need to run the appraisal program.

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 of each stratum in order to use this methodology.

### Input Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file

created with a spreadsheet package (e.g. Quattro Pro).

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

**7483 289.99 43.00**

Explanation:

- 7483 -** This is a number assigned by the user. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.
- 289.99 -** This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.
- 43.00 -** If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of \$289.99 claimed by a vendor, only \$246.99 was actually owed, then the difference amount entered would be \$43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

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

**9999 3E33**

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

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data

file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

**THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

**ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR DATAFILE?**

The user informs the program as to the format used in creating the data file of values. Instead of "**DATAFILE**", the actual name of the data file entered in the previous response by the user will be displayed.

**DO YOU WANT <C>OMPLETE, <P>ARTIAL OR <S>UMMARY APPRAISAL?**

Only the first letter of the appraisal options needs to be entered. The user may want to reduce printed output by having only a portion of the appraisal printed. The user may decide initially to appraise examined, audited and difference amounts for all strata (Complete). Subsequent appraisals of the sample after further review and modifications may be only for the difference amounts in selected strata (Partial) or only overall amounts (Summary).

**DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

### **NAME OF DISK FILE FOR APPRAISAL OUTPUT?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.APR."

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

While reading the data file, the program will display the name of the data file and the current record being processed. In displaying the current record, the program will limit the display to 50 characters.

After the program has read all of the values in the data file, a summary of the file will appear on the screen. If the user has selected the printer for output, the summary will also appear on the printer. The user should have summed the values in the sample prior to executing this program. At this point the user should reconcile the values to determine that the data file is complete and accurate.

### **CONTINUE PROCESSING WITH THE CURRENT FILE (Y/N)?**

By entering the first letter of the appropriate response, the user may continue with the appraisal or skip the appraisal portion of the program.

**IF YOU HAVE CREATED A FILE WITH THE  
UNIVERSE SIZES ENTER THE FILE NAME, ELSE HIT  
THE ENTER KEY?**

The user has the option of creating a file of universe sizes prior to executing the program or interactively entering the values during program execution. If a file name is entered, then the complete path must be entered for a universe file that does not exist in the same directory. The values in the universe file will be displayed on the screen for editing by the user. The format for this file is simply the universe sizes, in order by stratum, separated by one or more spaces. Commas are allowed in this data file.

If no file name is entered, then the program will assume that the user will enter the data interactively. For each entry, the user will be prompted with:

**ENTER THE UNIVERSE FOR STRATUM ##?** where ## is the current stratum number. This value may be entered with or without commas and will be used in estimating population characteristics.

**ENTER THE STRATUM NUMBER TO BE EDITED  
(ELSE HIT <ENTER>)?**

The user will be allowed to edit any universe sizes entered either from a data file or from the keyboard. If only the <Enter> key is hit, then the program will assume that no further editing is needed.

**ENTER THE NEW UNIVERSE FOR STRATUM ##?**

If the user had entered a stratum number to be edited, then this prompt will appear. When the user enters a universe size, the entry will be tested to ensure it is numeric and that the value is not less than the sample size. This value may be entered with commas.

**IF YOU WOULD CARE TO SAVE THIS SET OF  
UNIVERSE VALUES, ENTER A FILE NAME (ELSE  
HIT <ENTER>)?**

If any universe sizes were entered from the keyboard, then the user will have the option of saving the universe sizes as a file. If the file is to be saved in a directory other than the current directory, then a complete path must be given. For example, a file of universe values going to the "A" drive could be "A:UNIVS.DAT."

---

---

**ENTER THE STRATUM NUMBER OF A STRATUM TO  
BE PRINTED, ELSE HIT <ENTER>?**

If the user selected a Partial appraisal, then this prompt allows the user to enter a stratum number. The prompt will continue until all the strata desired are identified by the user. As each stratum is selected by the user, the stratum will be highlighted on the screen. If the user decides not to print a stratum already selected, then by entering the stratum number again the program will ask the user to confirm that the stratum is to be de-selected.

**ENTER APPRAISAL CHOICES: <E>XAMINED,  
<A>DJUSTED AND/OR <D>IFFERENCE (e.g. ED)**

If a Partial printout is selected and two values have been entered for each sample item, then the user will be prompted to indicate which values will be appraised and printed. The user must enter the first letter of each type of appraisal. The letters (i.e. EAD) may be entered in any order and with or without spaces between the letters.

An example of a query screen appears on page 4-21. The user entries are in **boldface**. The resultant output of the appraisal, including a summary of the data file, is shown on page 4-22.

### **Program Output**

For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed:

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

- OVERALL - STANDARD ERROR** - A measurement of the variation of the point estimate of the total with respect to all possible totals for this universe and these sample sizes.
- SKEWNESS** - A measure of the symmetry of the frequency distribution of the sample items. Accounting universes are usually right skewed (majority of items have a low value while a few items have a high value).
- KURTOSIS** - A measure of the peakedness or flatness of the frequency distribution of the sample values.
- POINT - ESTIMATE** - A single estimate for the universe total based on the sample mean multiplied by the universe size.
- CONFIDENCE - LEVEL** - The confidence 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 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.
- 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 -** This is the result of dividing the precision amount by the  
**PERCENT** point estimate and stating the result as a percentage.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES

DATE: 07/24/92 TIME:13:04  
 VARIABLES APPRAISAL  
 VERSION: MONTH YEAR

ENTER THE NAME OF YOUR DATA FILE OR QUIT? **A:\STRAT.DAT**

THE ALLOWABLE FILE FORMATS ARE:

1 - EXAMINED VALUES ONLY	4 - EXAMINED AND AUDITED VALUES
2 - AUDITED VALUES ONLY	5 - EXAMINED AND DIFFERENCE VALUES
3 - DIFFERENCE VALUES ONLY	6 - AUDITED AND DIFFERENCE VALUES
0 - EXIT THE PROGRAM	

ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR B:\AUDIT\OUTFILE1.NBR? **1**

DO YOU WANT <C>COMPLETE, <P>ARTIAL OR <S>UMMARY APPRAISAL? **C**

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ? **2**

=====NEW SCREEN=====

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES

DATE: 07/24/92 TIME:13:04  
 VARIABLES APPRAISAL

STRATUM NUMBER	SAMPLE SIZE	VALUE OF SAMPLE	NONZERO ITEMS
1	10	5,200.00	10
2	20	5,600.00	20
	30	10,800.00	30

CONTINUE PROCESSING WITH THIS CURRENT FILE (Y/N) ? **Y**

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES

DATE: 07/24/92 TIME:13:04  
 VARIABLES APPRAISAL

IF YOU HAVE CREATED A FILE WITH THE UNIVERSE SIZES  
 ENTER THE FILE NAME, ELSE HIT THE ENTER KEY?

STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE
1	<b>10,000</b>						

ENTER THE UNIVERSE FOR STRATUM 2? **2,000**  
 ENTER THE STRATUM NUMBER TO BE EDITED (ELSE HIT <ENTER>)?

IF YOU WOULD CARE TO SAVE THIS SET OF UNIVERSE VALUES, ENTER A FILE NAME  
 (ELSE HIT <ENTER>)?

ENTER THE STRATUM NUMBER OF A STRATUM TO BE PRINTED, ELSE HIT <ENTER>?

DEPARTMENT OF HEALTH & HUMAN SERVICES			
OIG - OFFICE OF AUDIT SERVICES			
VARIABLES APPRAISAL			
DATE: 07/24/92			TIME: 13:04
DATA FILE USED: A:\STRAT.DAT			
STRATUM NUMBER	SAMPLE SIZE	VALUE OF SAMPLE	NONZERO ITEMS
1	10	5,200.00	10
2	20	5,600.00	20
	30	10,800.00	30
=====NEW SCREEN=====			
-----E X A M I N E D-----			
STRATUM 1	MEAN / UNIVERSE	520.00	10,000
	STANDARD DEVIATION	293.32	
	STANDARD ERROR	92.71	
	SKEWNESS	0.00	
	KURTOSIS	2.43	
	POINT ESTIMATE	5,200	
		----90 PERCENT----	----95 PERCENT----
	LOWER LIMIT	3,500,653	3,102,934
	UPPER LIMIT	6,899,347	7,297,066
	PRECISION AMOUNT	1,699,347	2,097,066
	PRECISION PERCENT	32.68%	40.33%
OVERALL	POINT ESTIMATE / UNIVERSE	5,760,000	12,000
	STANDARD ERROR	927,687	
		----90 PERCENT----	----95 PERCENT----
	LOWER LIMIT	4,234,093	3,941,770
	UPPER LIMIT	7,285,907	7,578,230
	PRECISION AMOUNT	1,525,907	1,818,230
	PRECISION PERCENT	26.49%	31.57%

---

## 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 Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

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 number assigned by the user. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.
- 289.99** - This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.
- 43.00** - If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of \$289.99 claimed by a vendor, only \$246.99 was actually owed, then the difference amount entered would be \$43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollars signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

The user must enter all the sample values by primary unit. When all the sample values for a primary unit have been entered, then the following line must be entered:

**9999 3E33**

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

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

#### THE ALLOWABLE FILE FORMATS ARE:

- |                            |                                  |
|----------------------------|----------------------------------|
| 1 - EXAMINED VALUES ONLY   | 4 - EXAMINED & AUDITED VALUES    |
| 2 - AUDITED VALUES ONLY    | 5 - EXAMINED & DIFFERENCE VALUES |
| 3 - DIFFERENCE VALUES ONLY | 6 - AUDITED & DIFFERENCE VALUES  |
| 0 - EXIT THE PROGRAM       |                                  |

#### ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR DATAFILE?

The user selects the format that was used in creating the data file of values. Instead of "DATAFILE", the actual name of the data file in the previous response by the user will be displayed.

#### DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

#### NAME OF DISK FILE FOR APPRAISAL OUTPUT?

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "A:FILEONE.APR."

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

While reading the data file, the program will display the name of the file being read and the current record being read. In displaying the current record, the program will limit the display to 50 characters.

After the program has read all of the values in the data file, a summary of the file will appear on the screen. If the user has selected the printer for output, the summary will also appear on the printer. The user should have summed the values in the sample prior to executing this program. At this point the user should reconcile the values to determine that the data file is complete and accurate.

### **IF YOU HAVE CREATED A FILE WITH THE UNIVERSE SIZES ENTER THE FILE NAME, ELSE HIT THE ENTER KEY?**

The user has the option of creating a file of universe sizes prior to executing the program or interactively entering the values during program execution. If a file name is entered, then the complete path must be entered for a data file that does not exist in the same directory. The values in the data file will be displayed on the screen for editing by the user. The format for the data file is simply the universe sizes, in order by primary unit, separated by one or more spaces. Commas are allowed in this file. Only the universe sizes for the primary units that were sampled are to be entered at this time. The total number of secondary items for all primary units in the universe will be requested later.

If no file name is entered, then the program will assume that the user will enter the data interactively. For each entry, the user will be prompted with: **ENTER THE UNIVERSE FOR PRIMARY UNIT ##?** where **##** is the current primary unit. This value may be entered with or without commas and will be used in estimating population characteristics.

---

---

**ENTER THE PRIMARY UNIT NUMBER TO BE  
EDITED (ELSE HIT <ENTER>)?**

The user will be allowed to edit any universe size entered either from a data file or from the keyboard after all the sampled primary units have been entered. If only the <Enter> key is hit, then the program will assume that no further editing is needed.

**ENTER THE NEW UNIVERSE FOR PRIMARY UNIT  
##?**

If the user had entered a primary unit number to be edited, then this prompt will appear. If the user enters a universe size, the entry will be tested to ensure it is numeric and that the value is not less than the sample size. This value may be entered with commas.

**IF YOU WOULD CARE TO SAVE THIS SET OF  
UNIVERSE VALUES, ENTER A FILE NAME (ELSE  
HIT <ENTER>)?**

If any universe sizes were entered from the keyboard, then the user will have the option of saving the universe sizes as a data file. If the file is to be saved in a directory other than the current directory, then a complete path must be given. For example, a file of universe values going to the "A" drive could be "A:UNIVS.DAT."

**ENTER THE NUMBER OF PRIMARY UNITS IN THE  
ENTIRE UNIVERSE?**

The user enters the number of all primary units in the universe. This quantity may be entered with commas and the program will only accept a numeric entry.

**ENTER THE NUMBER OF SECONDARY UNITS IN  
THE ENTIRE UNIVERSE (ELSE 0)?**

The user enters the total number of secondary units for all of the primary units in the universe. This quantity may be entered with commas and the program will only accept a numeric entry. If the number of secondary units is not known, a zero is entered and the output will not display the quantity of secondary units in the universe.

An example of the query screens are on page 4-30 and 4-31. The user entries are in **boldface**. The resultant output of the appraisal, including a summary of the data file, is shown on page 4-32.

## Program Output

For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed:

- 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.
- 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** - This is 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** - This is 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 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** - This is the result of dividing the precision amount by the point estimate.
- Z-VALUE USED** - This value is referred to as the standard normal deviate. It is a measurement from the point estimate to a confidence limit measured in standard errors.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92      TWO STAGE VARIABLE APPRAISAL      TIME:16:37
                    VERSION: MONTH YEAR

ENTER THE NAME OF YOUR DATA FILE OR QUIT? B:MULTOTT.DAT

                THE ALLOWABLE FILE FORMATS ARE:
1 - EXAMINED VALUES ONLY      4 - EXAMINED AND AUDITED VALUES
2 - AUDITED VALUES ONLY      5 - EXAMINED AND DIFFERENCE VALUES
3 - DIFFERENCE VALUES ONLY    6 - AUDITED AND DIFFERENCE VALUES
0 - EXIT THE PROGRAM

ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR B:\MULTOTT.DAT? 1

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)? 1

=====NEW SCREEN=====

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92      TWO STAGE VARIABLE APPRAISAL      TIME: 16:37

PRIMARY      SAMPLE      ===== E X A M I N E D =====
UNIT         SIZE      VALUE IN SAMPLE      NONZERO ITEMS
1           10           54.00           9
2           13           52.00           12
3           9            51.00           8
4           10           48.00           8
5           10           43.00           9
6           12           46.00           10
7           8            40.00           8
8           13           50.00           12
9           8            39.00           8
10          11           55.00           10

TOTALS      104          478.00          94

CONTINUE PROCESSING WITH THIS CURRENT FILE (YES/NO)? Y

```

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
TWO STAGE VARIABLE APPRAISAL

DATE: 09/14/92

TIME:16:37

IF YOU HAVE CREATED A FILE WITH THE UNIVERSE SIZES  
ENTER THE FILE NAME, ELSE HIT THE ENTER KEY? **B:\MULTOTT.UNV**

PRIME UNIT	UNIVERSE SIZE						
1	<b>50</b>	2	<b>65</b>	3	<b>45</b>	4	<b>48</b>
5	<b>52</b>	6	<b>58</b>	7	<b>42</b>	8	<b>66</b>
9	<b>40</b>	10	<b>56</b>				

ENTER THE NUMBER OF PRIMARY UNITS IN THE ENTIRE UNIVERSE? **90**  
ENTER THE NUMBER OF SECONDARY UNITS IN THE ENTIRE UNIVERSE? **4,500**

DEPARTMENT OF HEALTH & HUMAN SERVICES					
OIG - OFFICE OF AUDIT SERVICES					
TWO STAGE VARIABLE APPRAISAL					
DATE: 09/14/92				TIME: 16:37	
=====E X A M I N E D=====					
UNIT NBR	SAMPLE SIZE	SAMPLE MEAN	VARIANCE	UNIVERSE SIZE	POINT ESTIMATE
1	10	5.40	11.38	50	270
2	13	4.00	10.67	65	260
3	9	5.67	16.75	45	255
4	10	4.80	13.29	48	230
5	10	4.30	11.12	52	224
6	12	3.83	14.88	58	222
7	8	5.00	5.14	42	210
8	13	3.85	4.31	66	254
9	8	4.88	6.13	40	195
10	11	5.00	11.80	56	280
	104	4.80		522	2,400
NOT SAMPLED:					
	80			3,978	
OVERALL TOTALS:					
	90			4,500	21,602
STANDARD ERROR					867
CONFIDENCE LEVEL				---90 PERCENT--	---95 PERCENT--
LOWER LIMIT				20,176	19,903
UPPER LIMIT				23,027	23,300
PRECISION AMOUNT				1,425	1,699
PRECISION PERCENT				6.60%	7.86%
Z-VALUE USED				1.64485	1.95996

---

## 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 Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

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 sampled third stage unit. Each data line consists of a line number for that sampled third stage 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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

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

```
REGION I 8 2
MAINE 100
1 120.34 85.50
```

Explanation:

- REGION I** - 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.
- 8** - This is the universe of secondary units that are within this primary unit.
- 2** - This is the number of secondary units that were sampled from this primary unit.
- 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 of third stage items within this secondary unit.
- 1** - This is a number assigned by the user to a sample item. For ease of reference in this example, the number will be referred to as the line number.
- 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 the line number and this numeric value.
- 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 may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

The user must enter all the sample values by secondary unit. When all the sample values for a secondary unit have been entered, then the following line must be entered:

**9999 3E33**

This line serves as a sentinel. The line number may be any number selected by the user.

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

#### **THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

#### **ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR DATAFILE?**

The user informs the program as to the format used in creating the data file of values. Instead of "**DATAFILE**", the actual name of the data file entered in the previous response by the user will be displayed.

**ENTER THE NUMBER OF PRIMARY UNITS IN THE UNIVERSE?**

The user needs to enter the quantity of primary units in the universe. The quantity may be entered with commas and the program will only accept a numeric entry.

**DO YOU WANT OUTPUT TO PROVIDE <C>OMplete OR  
<S>UMMARY INFORMATION?**

The user has the option of printing appraisal information for each secondary unit sampled (complete) or only information for each primary unit sampled (summary). The overall results will displayed for either option. Only the first letter (**C** or **S**) needs to be entered.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER  
OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**NAME OF DISK FILE FOR APPRAISAL OUTPUT?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILEONE.APR.**"

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE A:FILEONE.APR** to have the results displayed on the screen or **TYPE A:FILEONE.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

While reading the data file, the program will display the name of the data file being read and the current record being processed. In displaying the current record, the program will limit the display to 50 characters.

An example of the query screen appears on page 4-39. The user entries are in **boldface**. The resultant output of the appraisal, including a summary of the data file, is shown on page 4-40.

### 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:

**FIRST STAGE / -** The description of the first or second stage unit as  
**SECOND STAGE** shown in the data file.

**SAMPLE -** The number of items sampled in the particular second stage.  
**SIZE**

**NON -** The quantity of sample items in the particular second stage that had  
**ZEROES** values other than zero.

**SAMPLE -** The average value for the sample items appraised. It is  
**MEAN** 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 -** This is the quantity of the third stage items within each sampled second  
stage unit from which the sample was drawn.

**POINT -** This is the single estimate for the universe total of the  
**ESTIMATE** 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** - This is 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** - This is the result of dividing the precision amount by the point estimate.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
DATE: 09/14/92 THREE STAGE VARIABLE APPRAISAL TIME:16:37  
VERSION: MONTH YEAR

ENTER THE NAME OF YOUR DATA FILE OR QUIT? **B:3STAGE.DTA**

THE ALLOWABLE FILE FORMATS ARE:

1 - EXAMINED VALUES ONLY	4 - EXAMINED AND AUDITED VALUES
2 - AUDITED VALUES ONLY	
5 - EXAMINED AND DIFFERENCE VALUES	
3 - DIFFERENCE VALUES ONLY	6 - AUDITED AND DIFFERENCE VALUES
	0 - EXIT THE PROGRAM

ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR B:3STAGE.DTA? **1**

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)? **2**

DEPARTMENT OF HEALTH & HUMAN SERVICES						
OIG - OFFICE OF AUDIT SERVICES						
THREE STAGE VARIABLE APPRAISAL						
DATE: 09/14/92						TIME: 16:37
NAME OF INPUT FILE: 3STAGE.DTA						
===== E X A M I N E D =====						
FIRST STAGE	SAMPLE	NON-	SAMPLE MEAN	VARIANCE	UNIVERSE	POINT ESTIMATE
SECOND STAGE	SIZE	ZEROES				
=====	=====	=====	=====	=====	=====	=====
REGION I						
MAINE	10	10	109.00	36.67	100	10,900
VERMONT	10	10	129.00	36.67	120	15,480
COMBINED	20		13,190		2	26,380
REGION II						
NEW YORK	10	10	189.00	36.67	80	15,120
COMBINED	10		15,120		1	15,120
REGION VI						
TEXAS	10	10	269.00	36.67	90	24,210
ARKANSAS	10	10	289.00	36.67	110	31,790
COMBINED	20		28,000		2	56,000
	STAGES		UNIVERSE		SAMPLED	
	FIRST		15		3	
	SECOND		5		5	
	THIRD		500		50	
OVERALL POINT ESTIMATE					487,500	
OVERALL STANDARD ERROR					163,566	
CONFIDENCE LEVEL					---90 PERCENT--	---95 PERCENT--
LOWER LIMIT					218,458	166,917
UPPER LIMIT					756,542	808,083
PRECISION AMOUNT					269,042	320,583
PRECISION PERCENT					55.19%	65.76%

---

## 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 Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

Prior to executing this program the user must create a 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).

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

For each stratum the user needs to use the following format.

#### **PRIVATE UNIVERSITIES 400**

Explanation:

- PRIVATE - UNIVERSITIES** This is a description of the clusters in the stratum. 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.
- 400 -** This is the number of clusters in the universe for the stratum. The quantity must be separated from the stratum description by one or more

spaces. The number may contain commas.

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.

**GOOD UNIVERSITY 40 \$12,000**

Explanation:

**GOOD -  
UNIVERSITY**

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.

**40 -**

This is the number of items reviewed within the cluster. The quantity must be separated from the cluster description by one or more spaces. The number may contain commas.

**\$12,000 -**

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, decimal point, dollar sign or other similar symbols. If the value is negative, it must be preceded by a minus sign (-).

After all the cluster lines of data have been entered for a stratum, the user must enter the following line.

**3E33**

This line serves as a sentinel for the program. This line is optional after the last stratum in the file.

In responding to the name of the data file, the user must enter the complete path if the data is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data

file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

**DO YOU WANT OUTPUT TO PROVIDE <C>OMplete OR  
<S>UMMARY INFORMATION?**

The user has the option of displaying statistical information for each cluster by selecting the complete option or only the strata and overall statistics by selecting the summary option. The user only needs to enter the first letter of the option selected (i.e. **C** or **S**). If neither letter is entered, the program will continue to prompt the user for a selection.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER  
OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D**(disk), **P**(printer) or **S**(screen).

**ENTER NAME OF FILE TO RECORD THE RESULTS**

If the user selects the disk for the output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILEONE.OUT**."

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE FILEONE.OUT** to have the results displayed on the screen or **TYPE FILEONE.OUT > PRN** to have the results sent to the printer. The user could also retrieve the file through a word processing package such as WordPerfect.

## ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the query screen is on page 4-48. The user entries are in **boldface**. The resultant output of the appraisal is shown on page 4-48.

### 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 data file are displayed.

**SAMPLE - UNIVERSE**        This is the universe size of clusters for each stratum. The numbers were obtained from the data file.

**SAMPLE - SIZE**            This is 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 are displayed.

**SAMPLE - UNIVERSE**        This is the universe size for each cluster in a stratum. The numbers were obtained from the data file.

**SAMPLE - SIZE**            This is 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**        This is 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 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. If, however, the summary option was selected, then this information will appear on the same line as the stratum identification.

**SAMPLED - VALUE** - This is 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** - This is 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** - This line displays the total number of clusters in the universe and the actual number of clusters sampled.

**CLUSTER - UNIT TOTALS** - This line contains information for all clusters sampled. The information includes the total number of items reviewed and the value of those items for the characteristic being measured.

**OVERALL - POINT ESTIMATE** - This is 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** - This is 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** - 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).

**LOWER - LIMIT** - This is the lower bound of the confidence interval. It is calculated by subtracting the precision amount from the point estimate.

**UPPER -  
LIMIT** This is 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** This is the result of dividing the precision amount by the point estimate.

#### **DO YOU WANT THE DATA FILE PRINTED (Y/N)?**

After displaying the results, the program will allow the user to display the data file used in this appraisal. The user only needs to enter a **Y** or **N** to respond to this query.

#### **DO YOU WANT THE DATA FILE OUTPUT TO <P>RINTER OR <S>CREEN?**

If the appraisal results were not sent to a disk file, the user has the option of writing the data file to a printer or the screen. The user would respond by entering a **P** (printer) or **S** (screen).

#### **DO YOU WANT THE DATA FILE OUTPUT TO <D>ISK, <P>RINTER OR <S>CREEN?**

If, however, the user had selected a disk file for writing the appraisal, then in addition to the printer and screen the user has the option of appending the data file to the appraisal file. The user only needs to enter the first letter of the selected output device (i.e. **D**, **P** or **S**).

---

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selected **P** (printer) for output and did not select a printer for the appraisal output, then this query would be made. If the printer was selected in the appraisal queries earlier, then selecting a printer output for the data file would result in the data file being printed through the same port as previously identified.

Some computers have multiple parallel ports or have connections through a LAN to multiple ports. This query allows the user to select an appropriate port. The user only needs to enter the port number between 1 and 4 . If the output is going to LPT1, then the user has the option of only hitting the Enter key.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92          STRATIFIED CLUSTER VARIABLE APPRAISAL          TIME:16:37
                        VERSION: MONTH YEAR

ENTER THE NAME OF YOUR DATA FILE OR QUIT? A:STCLUS.DAT

DO YOU WANT OUTPUT TO PROVIDE <C>OMplete OR <S>UMMARY INFORMATION? S

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)? 1

=====NEW SCREEN=====

CONTINUE PROCESSING WITH THIS CURRENT FILE (YES/NO)? Y
    
```

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92          STRATIFIED CLUSTER VARIABLE APPRAISAL          TIME:16:37
                        NAME OF INPUT FILE: A:STCLUS.DAT
    
```

STRATUM IDENTIFICATION	SAMPLE UNIVERSE	SAMPLE SIZE	SAMPLED VALUE	POINT ESTIMATE
STATE SUPPORTED UNIVS	415	25	1,329,000.00	22,061,400
PRIVATE UNIVERSITIES	168	10	547,000.00	9,189,600
STRATA TOTALS	583	35		
CLUSTER UNIT TOTALS	200	200	1,876,000.00	
OVERALL POINT ESTIMATE				31,251,000
OVERALL STANDARD ERROR				2,418,432
CONFIDENCE LEVEL	--90 PERCENT--		--95 PERCENT--	
LOWER LIMIT	27,273,043		26,510,971	
UPPER LIMIT	35,228,957		35,991,029	
PRECISION AMOUNT	3,977,957		4,740,029	
PRECISION PERCENT	12.73%		15.17%	

---

---

## 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 Queries

**IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL,  
ENTER NAME OF FILE, ELSE HIT THE <ENTER> KEY?**

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 DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

For each stratum the user needs to use the following format.

**26,000 750**

Explanation:

**26,000** - This is the point estimate for one stratum. The program will allow commas to be included in the value.

**750** - This is the standard error for one stratum. The program will allow for commas to be included.

The results for each stratum should be on a separate line and the values separated by one or more spaces.

**NUMBER OF STRATA FOR THIS APPRAISAL?**

If the user selects to enter the data interactively, the program will prompt for the number of strata that will be appraised. The number entered must be between 2 and 20.

Once the number of strata is entered, column headings will appear and the user will be prompted to enter the point estimate and standard error for each stratum.

**ENTER STRATUM TO BE EDITED (ELSE HIT <ENTER>)?**

After the data for all strata have been entered, the program will allow the user to edit any of the values. At this prompt, the user will enter the appropriate number (between 1 and the total number of strata) to indicate the stratum that is to be edited. If the user only hits the Enter key, then the program assumes that the user has completed editing the data.

**ENTER THE POINT ESTIMATE?**

For the stratum identified in the prior query, the user may now enter a new point estimate. If the user does not want to change the point estimate, then the user only needs to hit the Enter key.

**ENTER THE STANDARD ERROR?**

For the stratum identified in the earlier query, the user may now enter a new standard error. If the user does not want to change the standard error, then the user only needs to hit the Enter key.

**IF YOU WANT TO SAVE THE SET OF VALUES, ENTER A FILE NAME (ELSE HIT <ENTER>?)**

If the data was entered interactively or any changes were made, then the user will be given the opportunity to save the information. The primary file name may be up to 8 characters. The user may also add an extension to the file with a period and up to 3 characters after the period. If the user wishes to place the file in a directory other than the current directory, then a complete path must be given to that directory. For example, if the data is to be stored on the "A" drive, the user could enter "A:STRDATA.DAT".

---

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER  
OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D**(disk), **P**(printer) or **S**(screen).

**ENTER NAME OF FILE TO RECORD THE RESULTS**

If the user selects the disk for the output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILEONE.OUT**."

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE FILEONE.OUT** to have the results displayed on the screen or **TYPE FILEONE.OUT > PRN** to have the results sent to the printer. The user could also retrieve the file through a word processing package such as WordPerfect.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for the output, the program will prompt for a parallel port number. If the output is to be sent to LPT1, the user has the option of entering a 1 or just hitting the Enter key. The program tests to see if the number is in the range 1 through 4. However, the program will not test to see if the port number entered is valid for the particular workstation.

An example of the query screen is on page 4-53. The user entries are in **boldface**. The resultant output of the appraisal is shown on page 4-53.

## 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** - This is 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** - This is 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** - 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).

**LOWER - LIMIT** - This is the lower bound of the confidence interval. It is calculated by subtracting the precision amount from the point estimate.

**UPPER - LIMIT** - This is the upper bound of the confidence interval. It is calculated by adding the precision amount to the point estimate.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92          STRATIFIED MULTISTAGE VARIABLE APPRAISAL          TIME:16:37
                        VERSION: MONTH YEAR

IF YOU HAVE CREATED A DATA FILE FOR THIS APPRAISAL, ENTER NAME OF FILE
ELSE HIT THE <ENTER> KEY?
NUMBER OF STRATA FOR THIS APPRAISAL  3

STRATUM   POINT ESTIMATE   STANDARD ERROR   STRATUM   POINT ESTIMATE   STANDARD ERROR
=====
    1         150,000         1,000           2         250,000         1,200
    3         300,000         1,500

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?  S
    
```

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 09/14/92          STRATIFIED MULTISTAGE VARIABLE APPRAISAL          TIME: 16:37

THE ESTIMATORS ARE BASED ON THE FOLLOWING ENTRIES:
STRATUM   POINT ESTIMATE   STANDARD ERROR
    1         150,000         1,000
    2         250,000         1,200
    3         300,000         1,500

===== RESULTS =====

POINT ESTIMATE          STANDARD ERROR
      700,000              2,166

90%          90%          95%          95%
LOWER LIMIT  UPPER LIMIT  LOWER LIMIT  UPPER LIMIT
696,438      703,562      695,755      704,245

(PRESS ANY KEY TO CONTINUE)
    
```

---

## RHC TWO-STAGE

### Purpose

This program performs a multistage variable appraisal of a Rao-Hartley-Cochran (RHC) sample. 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 Queries

#### **ENTER THE NAME OF YOUR DATA FILE OR QUIT.**

Prior to executing this program, the user must create a data file that contains identifying data and one or two pieces of information for each secondary item sampled. The first piece of information for each sampling unit is 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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

---

**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 may wish to continue on the same line adding more sample information. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

The user will separate primary units in the data file by entering a the following line at the end of each primary unit of values:

**9999 3E33**

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

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user would enter **A:FILENAME**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary

sample menu.

**THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

**ENTER THE APPROPRIATE NUMBER TO  
DESCRIBE THE FORMAT FOR DATAFILE?**

The user informs the program as to the format used in creating the data file of values. Instead of "**DATAFILE**", the actual name of the data file entered in the previous response by the user will be displayed.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER,  
OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer), or **S** (screen).

**NAME OF DISK FILE FOR APPRAISAL OUTPUT.**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "**A**" drive could be "**A:OUTDISK.APR**".

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE OUTDISK.APR** to have the results displayed on the screen or **TYPE OUTDISK.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

After the program has read all the values in the data file, a summary of the file will appear on the screen. If the user has selected the printer for output, the summary will

also appear on the printer. The user should have summed the values in the sample prior to executing this program. At this point the user should reconcile the values to determine that the data file is complete and accurate.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

**CONTINUE PROCESSING WITH THIS CURRENT FILE (Y/N)?**

By entering the first letter of the appropriate response (Yes or No), the user may continue with the appraisal or skip the appraisal portion of the program.

**ENTER THE FILE NAME CONTAINING THE PRIMARY UNIT INFORMATION, ELSE ENTER QUIT.**

When the user used the RHC sample selection program, a file was created containing information on the clustering of the primary units in the universe as well as the primary units selected. It is that file that this query is requesting. As in the previous filename query, the user must enter the complete path if the data file is not in the same directory as the statistical package (e.g. A:OUTDATA.PU). In response to this query, the program prints the contents of this file with corresponding headings.

If the user can not find the file, a file containing the following data must be created before the appraisal can be performed. The file must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

**PRIMARY #1 200 1,100 12,500 8**

Explanation:

**PRIMARY #1** - This is a description of a primary unit in the universe. A maximum of 30 characters, including spaces, will be accepted. One or more spaces must separate this entry from the next entry on the line.

**200** - This is the quantity of secondary units in the particular primary unit. The user may include commas in the value. One or more spaces must separate this entry from the next entry on the line.

**1,100** - This is the size factor assigned by the user to the primary unit. Commas and decimal points may be included in the values. One or more spaces must separate this entry from the next entry on the line.

**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. One or more spaces must separate this entry from the next entry on the line.

**8** - This is the quantity of primary units that were randomly placed into this group. Commas may be included in the value.

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.

**CONTINUE PROCESSING WITH THIS FILE (Y/N)?**

By entering the first letter of the appropriate response (Yes or No), the user may continue with the appraisal or exit the program.

An example of the query screens appears on page 4-62. The user entries are in **boldface**. The resultant output of the appraisal is shown on page 4-63.

## Program Output

The output from this program consists of the following.

**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 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 the 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 is 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 primary sample means multiplied by their respective secondary universes.

**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 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 (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.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
TWO STAGE RHC PROCEDURE
VERSION: MONTH YEAR
DATE: 06/01/92
TIME: 13:22

ENTER THE NAME OF YOUR DATA FILE OR QUIT. A:DATA2.RHC

THE ALLOWABLE FILE FORMATS ARE:
1 - EXAMINED VALUES ONLY          4 - EXAMINED & AUDITED VALUES
2 - AUDITED VALUES ONLY          5 - EXAMINED & DIFFERENCE VALUES
3 - DIFFERENCE VALUES ONLY       6 - AUDITED & DIFFERENCE VALUES
0 - EXIT THE PROGRAM

ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR A:DATA2.RHC? 1

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER, OR <S>CREEN? S
NAME OF DISK FILE FOR APPRAISAL OUTPUT. A:OUTDISK

===== NEW SCREEN =====

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT
TWO STAGE RHC PROCEDURE
DATE: 06/01/92
TIME: 13:22
DATA FILE USED: A:DATA2.RHC

PRIMARY UNIT      SAMPLE SIZE      ===EXAMINED===      NUMBER OF
=====          =====          =====          NONZERO ITEMS
=====          =====          =====          =====
1                2                830.00              2
2                2                722.00              2
3                2                576.00              2

TOTALS          6                2,128.00            6

CONTINUE PROCESSING WITH THIS CURRENT FILE (Y/N) ? Y

===== NEW SCREEN =====

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
TWO STAGE RHC PROCEDURE
DATE: 06/01/92
TIME: 13:22

ENTER THE FILE NAME CONTAINING THE PRIMARY UNIT INFORMATION, ELSE ENTER QUIT. A:OUTFILE
FILE OF PRIMARY UNIT INFORMATION: A:OUTFILE

P.U. NBR      PRIMARY UNIT ID      SECONDARY UNIVERSE      PRIMARY UNIT SIZE      GROUP SIZE      UNITS IN GROUP
=====          =====          =====          =====          =====          =====
1      ILLINOIS          16                7,521          42,379          6
2      NEW YORK          18                7,393          42,174          6
3      OHIO              16                6,131          46,372          7

TOTALS          50                21,045          130,925          19

CONTINUE PROCESSING WITH THIS FILE (Y/N) ? Y
    
```

DEPARTMENT OF HEALTH & HUMAN SERVICES					
OIG - OFFICE OF AUDIT SERVICES					
TWO STAGE RHC PROCEDURE					
DATE: 06/01/92					TIME: 13:22
DATA FILE USED: A:DATA2.RHC					
--- POINT ESTIMATES ---					
P.U. NBR	SAMPLE SIZE	SAMPLE MEAN	===EXAMINED=== SECONDARY UNIVERSE	SIZES RATIO	POINT ESTIMATE
=====	=====	=====	=====	=====	=====
1	2	415.00	16	5.635	37,415
2	2	361.00	18	5.705	37,068
3	2	288.00	16	7.564	34,853
TOTALS:	6		50		109,336
--- VARIANCE COMPONENTS ---					
P.U. NBR	WITHIN VARIANCE	BETWEEN VARIANCE	TOTAL VARIANCE		
1	9,119,289	5,378,386	14,497,675		
2	7,821,945	4,509,410	12,331,355		
3	433,723	17,996,058	18,429,781		
TOTALS:	17,374,957	27,883,854	45,258,811		
PRIMARY UNITS SAMPLED:					3
PRIMARY UNITS NOT SAMPLED:					16
PRIMARY UNITS IN POPULATION:					19
POINT ESTIMATE OF POPULATION TOTAL:					109,336
STANDARD ERROR				6,727	
CONFIDENCE LEVEL		---90 PERCENT--		---95 PERCENT--	
LOWER LIMIT		98,270		96,150	
UPPER LIMIT		120,402		122,521	
PRECISION AMOUNT		11,066		13,186	
PRECISION PERCENT		10.12%		12.06%	
Z-VALUE USED		1.644850		1.959960	

---

## RHC THREE-STAGE

### Purpose

This program performs a multistage variable appraisal of a Rao-Hartley-Cochran (RHC) sample. 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 Queries

#### **ENTER THE NAME OF YOUR DATA FILE OR QUIT.**

Prior to executing this program, the user must create a data file that contains identifying data and one or two pieces of information for each third stage item sampled. The first piece of information for each sampling unit is a numeric value the user wants to appraise (e.g. examined, audited, difference). 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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

---

**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 may wish to continue on the same line adding more sample information. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

The user will separate secondary units in the data file by entering a the following line at the end of each secondary unit of values:

**9999 3E33**

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

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:FILENAME**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary

sample menu.

**THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

**ENTER THE APPROPRIATE NUMBER TO  
DESCRIBE THE FORMAT FOR DATAFILE?**

The user informs the program as to the format used in creating the data file of values. Instead of "**DATAFILE**", the actual name of the data file entered in the previous response by the user will be displayed.

**ENTER THE FILE NAME CONTAINING THE PRIMARY AND  
SECONDARY INFORMATION, ELSE ENTER 'QUIT'?**

This file contains information about the primary and secondary units that were selected. The information must be aligned in a set format. The universe file is in order to correspond to the data file previously discussed. The primary universe information is entered first, followed by the data for the secondary universes. The format is as follows:

<b>PRIMARY #1</b>	<b>8</b>	<b>6,000</b>	<b>18,000</b>	<b>6</b>	<b>3</b>
<b>SECONDARY #1</b>	<b>3,500</b>	<b>1,000</b>	<b>1,400</b>	<b>10</b>	

Explanation:

- PRIMARY - #1** - This is a brief description of the primary unit that was sampled. The description should be no longer than 30 characters in length. One or more spaces must separate this entry from the next entry on the line.
- 8 -** This is the quantity of secondary units in this primary unit. The user may include commas in the value. One or more spaces must separate this entry from the next entry on the line.
- 6,000 -** This is the size factor assigned by the user to the primary unit. Commas and decimal points may be included in the values. One or

more spaces must separate this entry from the next entry on the line.

- 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. One or more spaces must separate this entry from the next entry on the line.
- 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. One or more spaces must separate this entry from the next entry on the line.
- 3** - This is the number of secondary units sampled from this primary unit for this appraisal. Commas may be used in the value.
- SECONDARY #1** - This is a brief description of the secondary unit that is sampled. A maximum of 30 characters, including spaces, will be accepted. One or more spaces must separate this entry from the next entry on the line.
- 3,500** - This is the quantity of third stage units in this secondary unit. The entry may contain commas and a decimal point. At least one space must separate this entry from the next entry on the line.
- 1,000** - This is the secondary size factor used in weighting the secondary unit. There must be at least one space separating this entry from the next entry on the line. 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. There must be at least one space separating this entry from the next entry on the line. 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.

**DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER,  
OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer), or **S** (screen).

### NAME OF DISK FILE FOR APPRAISAL OUTPUT.

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:OUTDISK.APR**".

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE OUTDISK.APR** to have the results displayed on the screen or **TYPE OUTDISK.APR > PRN** to have the results sent to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

After the program has read all the values in the data file, a summary of the file will appear on the screen. If the user has selected the printer for output, the summary will also appear on the printer. The user should have summed the values in the sample prior to executing this program. At this point the user should reconcile the values to determine that the data file is complete and accurate.

### ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the query screens appears on page 4-72. The user entries are in **boldface**. The resultant output of the appraisal is shown on pages 4-72 through 4-74.

## Program Output

The output from this program comes in three sections. The first section provides summary information from the 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 that were 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.

**NON-ZERO - COUNT** - The number of third stage sample items that had a value of 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 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.

**OVERALL SECTION:**

The results of the above information is 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 the summation of the primary sample means multiplied by their respective secondary universes.
- 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 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 (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.



DEPARTMENT OF HEALTH & HUMAN SERVICES OIG - OFFICE OF AUDIT SERVICES THREE STAGE RHC PROCEDURE			
DATE: 06/01/92			TIME: 13:22
--- POINT ESTIMATES --- *****E X A M I N E D*****			
**** SAMPLED UNITS ****	SIZES		POINT
PRIMARY / SECONDARY IDENTIFICATION	SAMPLE MEAN	RATIO	ESTIMATE
=====	=====	=====	=====
PRIMARY UNIT #10			
SECONDARY UNIT #1	6.50	1.4000	31,850
SECONDARY UNIT #7	9.50	3.6000	37,620
SECONDARY UNIT #6	12.50	2.5455	143,182
TOTAL			212,652
PRIMARY UNIT #9			
SECONDARY UNIT #4	15.50	2.5000	77,500
SECONDARY UNIT #3	18.50	1.7500	38,850
SECONDARY UNIT #9	21.50	6.0000	77,400
TOTAL			193,750
PRIMARY UNIT #1			
SECONDARY UNIT #1	24.50	1.6000	39,200
SECONDARY UNIT #6	27.50	2.8000	92,400
SECONDARY UNIT #10	30.50	2.5714	196,071
TOTAL			327,671
--- VARIANCE COMPONENTS FOR PRIMARY UNITS ---			
**** SAMPLED UNITS ****	WITHIN	BETWEEN	TOTAL
PRIMARY UNIT IDENTIFICATION	VARIANCE	VARIANCE	VARIANCE
=====	=====	=====	=====
PRIMARY UNIT #10	66,777,617	2,591,687,449	2,658,465,066
PRIMARY UNIT #9	13,358,583	150,946,970	164,305,553
PRIMARY UNIT #1	19,790,414	3,687,039,035	3,706,829,450
--- COMBINED VARIANCE COMPONENTS ---			
STAGE 1	STAGES 2 AND 3	TOTAL VARIANCE	
=====	=====	=====	
372,682,471,167	25,231,197,156	397,913,668,324	

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 06/01/92	THREE STAGE RHC PROCEDURE	TIME: 13:22
*****E X A M I N E D*****		
--- SUMMARY OF APPRAISAL RESULTS ---		
PRIMARY UNITS SAMPLED	3	
PRIMARY UNITS NOT SAMPLED	7	
TOTAL PRIMARY UNITS	10	
OVERALL POINT ESTIMATE	2,790,602	
OVERALL STANDARD ERROR	630,804	
CONFIDENCE LEVEL	---90 PERCENT--	---95 PERCENT--
LOWER LIMIT	1,753,024	1,554,251
UPPER LIMIT	3,828,180	4,026,952
PRECISION AMOUNT	1,037,578	1,236,351
PRECISION PERCENT	37.18%	44.30%
DATA FILE USED FOR SAMPLE RESULTS:	A:DATA3.RHC	
FILE OF RANDOM GROUPINGS:	A:OUT3.RHC	

## POST STRATIFICATION

### 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 Queries

#### ENTER THE NAME OF YOUR DATA FILE OR QUIT?

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.

The data file containing the above information must be stored in a DOS (ASCII) 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. WordPerfect) or a print file created with a spreadsheet package (e.g. Quattro Pro).

The format for the data file is as follows:

**7483 289.99 43.00**

Explanation:

- 7483** - is a number assigned by the user. If the user is creating the data file with BASIC, then this number would be the line number. The user should use the sample item number as the number in this position. For ease of reference in this example, the number will be referred to as the line number.
- 289.99** - This is a number being reviewed by the user. The number, for example, could be a dollar amount claimed or the number of items on an inventory card. If the number is negative, then a minus sign must precede the number. The user must insert at least one space between the line number and the first numeric value entered on the line.
- 43.00** - If two pieces of information are being gathered for each sampling unit, then this number is the audited or difference amount determined by the user. For example, if the user had determined that of \$289.99 claimed by a vendor, only \$246.99 was actually owed, then the difference amount entered would be \$43.00.

The user may wish to continue on the same line adding more sample information. The line number appears only once on each line. Each piece of information must be separated by at least one space on the line. For ease of editing, the user should enter the data for each sampling unit on a separate line. Enough spaces should be placed between values so that the sample values align vertically. The data values may contain commas and dollar signs (\$). The program assumes one or more spaces as the only delimiter between pieces of data.

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

**9999 3E33**

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

In responding to the name of the data file, the user must enter the complete path if the data file is not in the same directory as the statistical package. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:DATAFILE.DAT**.

If the program cannot locate the file, the user will be prompted to re-enter the path and file name. The user has the option of entering **QUIT** when prompted for a data file name. **QUIT** causes the program to terminate and return the user to the primary sample menu.

**THE ALLOWABLE FILE FORMATS ARE:**

- |                                   |   |
|-----------------------------------|---|
| <b>1 - EXAMINED VALUES ONLY</b>   | <b>4 - EXAMINED &amp; AUDITED VALUES</b>    |
| <b>2 - AUDITED VALUES ONLY</b>    | <b>5 - EXAMINED &amp; DIFFERENCE VALUES</b> |
| <b>3 - DIFFERENCE VALUES ONLY</b> | <b>6 - AUDITED &amp; DIFFERENCE VALUES</b>  |
| <b>0 - EXIT THE PROGRAM</b>       |   |

**ENTER THE APPROPRIATE NUMBER TO DESCRIBE THE FORMAT FOR DATAFILE?**

The user informs the program as to the format used in creating the data file of values. Instead of "**DATAFILE**", the actual name of the data file entered in the previous response by the user will be displayed.

**DO YOU WANT <C>OMplete, <P>ARTIAL OR <S>UMMARY APPRAISAL?**

Only the first letter of the appraisal options needs to be entered. The user may want to reduce printed output by having only a portion of the appraisal printed. The user may decide initially to appraise examined, audited and difference amounts for all strata (Complete). Subsequent appraisals of the sample after further review and modifications may be only for the difference amounts in selected strata (Partial) or only overall amounts (Summary).

**OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**ENTER THE OUTPUT FILE NAME FOR THIS APPRAISAL?**

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "**A**" drive would be "**A:FILEONE.APR.**"

To print the disk file, the user could enter **TYPE FILENAME** to have the results displayed on the computer or **TYPE FILENAME > PRN** to have results output to a printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects a printer for the output, then this query will appear. Computers that use multiple parallel ports or have connections through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

While reading the data file, the program will display the name of the file being read and the current record being read. In displaying the current record, the program will limit the display to 50 characters.

After reading the data file, the program will proceed with the following questions for the user.

### **IF YOU HAVE CREATED A FILE WITH THE UNIVERSE SIZES ENTER THE FILE NAME, ELSE HIT THE ENTER KEY?**

The user has the option of creating a file of universe values prior to executing the program or interactively entering the values during program execution. If a file name is entered, then the complete path must be entered for a data file that does not exist in the current directory. The values in the data file will be displayed on the screen for editing by the user. The format for the data file is simply the universe sizes, in order by stratum, separated by one or more spaces. Commas, dollar signs and other symbols are allowed in the data file.

If no file name is entered, then the program will assume that the user will enter the data interactively. For each entry the user will be prompted with:

### **ENTER THE UNIVERSE FOR STRATUM ##?**

The ## represents the stratum number that will actually appear at each prompt. The value entered will be tested to ensure that it is numeric and at least equal to the sample size for the stratum. If it is less than the sample size, the user will be asked to re-enter the universe value. The value may be entered with commas, dollar signs or other symbols.

---

**ENTER THE STRATUM NUMBER TO BE EDITED (ELSE HIT <ENTER>)?**

The user will be allowed to edit any universe sizes entered either from a data file or from the keyboard. If only the Enter key is hit, then the program will assume that no further editing is needed.

**ENTER THE NEW UNIVERSE FOR STRATUM ##?**

If the user had entered a stratum number to be edited, then this prompt will appear. If the user enters a value, the entry will be tested to ensure it is numeric and that the value is at least equal to the sample size. The value may be entered with commas, dollar signs or other symbols.

**IF YOU WOULD CARE TO SAVE THIS SET OF UNIVERSE VALUES, ENTER A FILE NAME (ELSE HIT <ENTER>)?**

If any universe sizes were entered from the keyboard, then the user will have the option of saving the universe values as a data file. If the file is to be saved in a directory other than the current directory, then a complete path must be given. For example, a file of universe values going to the "A" drive would be "A:UNIVS.DAT."

**ENTER THE STRATUM NUMBER OF A STRATUM TO BE PRINTED, ELSE HIT <ENTER>?**

If the user selected a Partial appraisal, then this prompt allows the user to enter a stratum number. The prompt will continue until all the strata desired are identified by the user. As each stratum is selected by the user, the stratum will be highlighted on the screen. If the user decides not to print a stratum already selected, then by entering the stratum number again the program will ask the user to confirm that the stratum is to be de-selected.

**ENTER APPRAISAL CHOICES: <E>XAMINED, <A>DJUSTED AND/OR <D>IFFERENCE (e.g. ED)**

If a partial printout is selected and two values have been entered for each sample item, then the user will be prompted to indicate which values will be appraised and printed. The user must enter the first letter of each type of appraisal. The letters (i.e. EAD) may be entered in any order and with or without spaces between the letters.

An example of a query screen is on page 4-82. The user entries are in **boldface**. The resultant output of the appraisal is shown on page 4-83.

### Program Output

For the examined, adjusted and difference sections of the output, the following pieces of information will be displayed for each stratum.

**SAMPLE SIZE** - The number of items sampled in the particular stratum.

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

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

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

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

The summary section will display the following information.

**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.

The overall limits and precision are presented for both the 90 and 95 percent (two-sided) confidence levels.

After the results have been displayed, the user will be asked the following questions.

**DO YOU WANT YOUR DATA FILE PRINTED (Y/N)?**

This option allows the user to have the data file sent to a printer. The user must respond with a Y or a N in order to proceed.

**ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user responds with a **Y** to the previous question, then this query will appear. Computers that use multiple parallel ports or have connections through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
POST STRATIFICATION APPRAISAL  
VERSION: MONTH YEAR

DATE: 06/26/92 TIME: 12:00

ENTER THE NAME OF YOUR DATA FILE OR QUIT? **B:OTT.DAT**

THE ALLOWABLE FILE FORMATS ARE:

1 - EXAMINED VALUES ONLY                      4 - EXAMINED AND AUDITED VALUES  
2 - AUDITED VALUES ONLY  
5 - EXAMINED AND DIFFERENCE VALUES  
3 - DIFFERENCE VALUES ONLY  
6 - AUDITED AND DIFFERENCE VALUES  
0 - EXIT THE PROGRAM

ENTER THE APPROPRIATE FORMAT NUMBER FOR B:OTT.DAT? **1**

DO YOU WANT <C>OMplete, <P>ARTIAL OR <S>UMMARY APPRAISAL? **C**

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? **P**

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ?

===== NEW SCREEN =====

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
POST STRATIFICATION APPRAISAL  
VERSION JUNE 1992

DATE: 06/26/92 TIME: 12:00

IF YOU HAVE CREATED A FILE WITH THE UNIVERSE SIZES ENTER THE FILE NAME, ELSE  
HIT THE ENTER KEY?

STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE	STRT. NBR.	UNIVERSE SIZE
<b>1</b>	<b>4,000</b>	<b>2</b>	<b>550</b>				

ENTER STRATUM NUMBER TO BE EDITED (ELSE HIT <ENTER>)?

DEPARTMENT OF HEALTH & HUMAN SERVICES			
OIG - OFFICE OF AUDIT SERVICES			
POST STRATIFICATION APPRAISAL			
DATE: 06/26/92			TIME: 12:00
DATA FILE USED: B:OTT.DAT			
STRATUM 1	SAMPLE SIZE / UNIVERSE	70	4,000
	MEAN	520.00	
	STANDARD DEVIATION	209.96	
	POINT ESTIMATE	2,080,000	
STRATUM 2	SAMPLE SIZE / UNIVERSE	30	550
	MEAN	280.00	
	STANDARD DEVIATION	90.05	
	POINT ESTIMATE	154,000	
OVERALL	SAMPLE SIZE / UNIVERSE	100	4,550
	POINT ESTIMATE	2,234,000	
	STANDARD ERROR	89,838	
		=====	=====
		90 PERCENT	95 PERCENT
	LOWER LIMIT	2,086,230	2,057,922
	UPPER LIMIT	2,381,770	2,410,078
	PRECISION AMOUNT	147,770	176,078
	PRECISION PERCENT	6.61%	7.88%

---

---

## UNKNOWN UNIVERSE

### 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 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 Queries

The program will first issue a warning to the user that prior appraisals of the two samples must be taken before running this module. After issuing the warning, the user will be asked:

#### **DO YOU WANT TO CONTINUE WITH THIS PROGRAM (Y/N)?**

The user will enter a **Y** (YES) in order to proceed with the program or a **N** (NO) to return to the main menu.

For the sample used to estimate the population, the program will ask:

#### **UNIVERSE FROM WHICH ITEMS WERE SAMPLED?**

The population 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 TAKEN?**

The user enters the number of sampling units (e.g. file drawers) drawn from the universe entered above.

**MEAN OF THE SAMPLE TAKEN?**

The variable appraisal program will generate a mean for the sample. The user should now enter that mean value.

**STANDARD DEVIATION OF THE SAMPLE TAKEN?**

The variable appraisal program will generate a standard deviation for the sample. The user should now enter that standard deviation value.

The following queries relate to the sample drawn to evaluate a set of criteria:

**SAMPLE SIZE TAKEN?**

The user has drawn a second sample of items that meet the criteria for review. The user should now enter the size of this second sample.

**MEAN OF THE SAMPLE TAKEN?**

The variable appraisal program previously used with this sample generated a sample mean. The user should now enter that mean value.

**STANDARD DEVIATION OF THE SAMPLE TAKEN?**

The variable appraisal program previously used with this sample generated a sample standard deviation. The user should now enter that standard deviation value.

**OUTPUT APPRAISAL TO <D>ISK, <P>RINTER OR <S>CREEN?**

The user selects the appropriate output by responding to the query with a **D** (disk), **P** (printer) or **S** (screen).

**ENTER THE OUTPUT FILE NAME FOR THIS APPRAISAL?**

By sending the output to disk, the user can then transmit the information by telecommunications to another microcomputer. To print the disk file, the user could enter **TYPE FILENAME** to have the results displayed on the computer or **TYPE FILENAME > PRN** to have results output to a printer. The user could also retrieve the file with a word processing package such as WordPerfect.

If the user selects the disk for output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a disk file going to the "A" drive could be "**A:FILEONE.APR.**"

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects a printer for the output, then this query will appear. Computers that use multiple parallel ports or have connections through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

Examples of the warning and query screens appear on page 4-88. The user entries are in **boldface**. The resultant output, including a summary of the data entered, is shown on page 4-89.

### **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** - Universe from which the sample was drawn to estimate the population.

**SAMPLE** - The two sample sizes that were drawn for this evaluation.

**MEAN** - The two mean values entered by the user.

**STANDARD DEVIATION** - The two values entered by the user.

The second section of the output displays the results of the estimation.

**POINT ESTIMATE** - This is the estimate for the 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. For the examined (reviewed) appraisal, the universe total may be known and should be reviewed by the user to see if, in fact, the actual value does fall within the confidence interval.

**PRECISION -  
PERCENT**        This is the result of dividing the precision amount by the point estimate.

The precision information is given at both the 90 and 95 percent (two-sided) confidence levels.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 DATE: 06/26/92 VARIABLE APPRAISAL WITH UNKNOWN POPULATION TIME: 12:17  
 VERSION: MONTH YEAR

**WARNING!!!**

THIS PROGRAM CALCULATES THE OVERALL PRECISION WHEN THE POPULATION SIZE IS UNKNOWN FOR A VARIABLE PROJECTION. A VARIABLE APPRAISAL SHOULD BE RUN FOR EACH OF THE SAMPLES PRIOR TO EXECUTING THIS PROGRAM. THIS PROGRAM WILL ASK FOR THE MEAN AND STANDARD DEVIATION FROM EACH OF THE TWO SAMPLE APPRAISALS.

DO YOU WANT TO CONTINUE WITH THIS PROGRAM (Y/N) ?

DEPARTMENT OF HEALTH & HUMAN SERVICES  
 OIG - OFFICE OF AUDIT SERVICES  
 DATE: 06/26/92 VARIABLE APPRAISAL WITH UNKNOWN POPULATION TIME: 12:17  
 VERSION: MONTH YEAR

FOLLOWING QUERIES RELATE TO THE SAMPLE USED TO ESTIMATE THE POPULATION:

UNIVERSE FROM WHICH ITEMS WERE SAMPLED?	575
SAMPLE SIZE TAKEN?	70
MEAN OF THE SAMPLE TAKEN?	10.00
STANDARD DEVIATION OF THE SAMPLE TAKEN?	2.00

FOLLOWING QUERIES RELATE TO THE SAMPLE USED TO ESTIMATE A VARIABLE IN THE POPULATION:

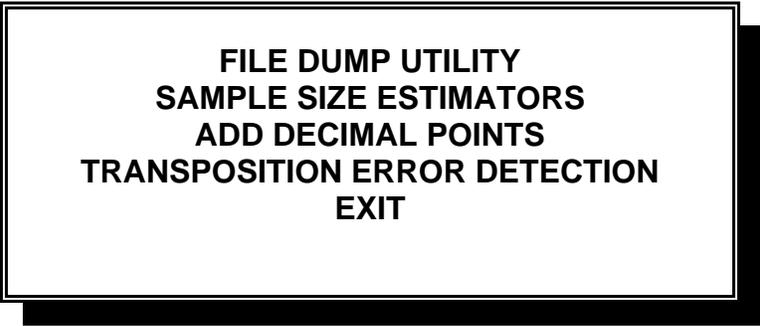
SAMPLE SIZE TAKEN?	55
MEAN OF THE SAMPLE TAKEN?	15.00
STANDARD DEVIATION OF THE SAMPLE TAKEN?	5.00

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER, OR <S>CREEN? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ?

DEPARTMENT OF HEALTH & HUMAN SERVICES		
OIG - OFFICE OF AUDIT SERVICES		
DATE: 06/26/92	VARIABLE APPRAISAL WITH UNKNOWN POPULATION	TIME: 12:17
===== GIVEN =====		
	SAMPLE TO ESTIMATE POPULATION	SAMPLE FOR VARIABLE
UNIVERSE	575	
SAMPLE	70	55
MEAN	10.00	15.00
STANDARD DEVIATION	2.00	5.00
===== ESTIMATION =====		
	90% CONFIDENCE	95% CONFIDENCE
POINT ESTIMATE	86,250	86,250
LOWER LIMIT	79,154	77,795
UPPER LIMIT	93,346	94,705
PRECISION AMOUNT	7,096	8,455
PRECISION PERCENT	8.23%	9.80%

## OVERVIEW



**FILE DUMP UTILITY  
SAMPLE SIZE ESTIMATORS  
ADD DECIMAL POINTS  
TRANSPOSITION ERROR DETECTION  
EXIT**

In the course of working with data for statistical samples, we have found the need to develop certain utility programs to assist us and the staff. The following programs may be of some benefit to users of this package. A brief description of each program is given below. A detailed explanation of how to use each module is described later in this section.

### **FILE DUMP UTILITY**

When data is downloaded from a mainframe computer or developed by others, unprintable characters (e.g. nulls, tabs, etc.) may not be readily detected by a user. By obtaining a decimal dump of a few records in a file, a user may quickly determine the cause of data processing errors. The program presents an ASCII translation and the decimal equivalent of each byte displayed.

### **SAMPLE SIZE ESTIMATORS**

A reviewer may want to determine the optimum sample sizes for a given universe. This utility allows such estimates for a unrestricted or stratified variable sample. This program will also allow a user to determine the optimum distribution of a sample among strata when the overall sample size has been determined.

## **ADD DECIMAL POINTS**

It is common for numeric data on a mainframe computer to be stored without decimal points. When the data is downloaded to a microcomputer, the decimal points may still be missing. This program allows the user to indicate where the decimal points should be added and then creates a new file with the decimal points included.

## **TRANSPOSITION ERROR DETECTION**

In transcribing large quantities of numeric data, it is not uncommon for transposition errors to occur. When the user enters the amount of difference between the original column of numeric data and the transcribed column, the program will determine if it might be a transposition error and the possible combinations that would cause such a difference.

---

---

## FILE DUMP UTILITY

### Purpose

This program allows the user to obtain an ASCII decimal dump of all or a part of any file. The utility allows the user to identify the location of certain data items within a record as well as the disclosure of unprintable characters that might be causing problems in a software application. The output is similar to mainframe dump utilities.

### Input Queries

#### NAME OF FILE TO BE DISPLAYED?

The user enters the name of the file that is to be dumped. If the file is not in the current directory, the complete path to the file must be given. An error message will be given if the file can not be located.

#### DO YOU WANT THE DUMP SENT TO THE <S>CREEN OR <P>RINTER?

The user selects the location of the output by entering a **S** (Screen) or **P** (Printer).

#### ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

#### HOW MANY CHARACTERS (BYTES) DO YOU WANT DUMPED?

The user enters in the number of characters to be displayed. The dump will begin with the first character (byte) and continue to display characters until the quantity entered is reached. If the quantity entered is greater than the length of the file, the program will only display the entire contents of the file with no error message. The quantity entered may contain commas. The largest number that may be entered is 32,767.

An example of the input screen is shown on page 5-5. The user entries are in **boldface**. The requested dump is also shown on page 5-5.

### Program Output

The output from the program is read in columnar format. There are four rows of information printed for each character. The top row contains the ASCII translation of the byte. If the ASCII decimal value is between 32 and 126, the ASCII character will be printed. If the decimal value is outside of this range, then a period (.) will be printed in the top row. The next three rows contain the decimal equivalent of the character. The value is read vertically. The following is an example of a one character display.

```
#  
0  
3  
5
```

Explanation:

# - is the ASCII character  
035 - is the decimal equivalent of the character.

The output is displayed in 50 character segments. The range of each segment is shown to the left of the first row of the segment. Across the top of each page are columnar headings to assist in identifying a particular location. The last digit of each number printed in the columnar heading is the actual location for that number.



---

---

## SAMPLE SIZE ESTIMATORS

### Purpose

This program allows the user to estimate the sample size for a certain precision at a given confidence level. The sample size estimates are based on the assumption that a variable appraisal will be performed.

The program will generate sample sizes for unrestricted and stratified samples. The user may also enter the overall sample size for a stratified sample and the program will determine the optimum allocation among the strata. The user needs to enter the mean, universe and standard deviation for each stratum in order to generate the estimated sample sizes. The program allows for sample sizes for up to 12 strata.

### Input Queries

#### ENTER THE NAME OF THIS AUDIT/REVIEW?

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

#### ENTER THE TOTAL SAMPLE SIZE, ELSE 0 (ZERO)?

If the user plans on performing a stratified sample and the overall sample size has been predetermined, then the overall sample size should be entered. If the overall sample size has not been predetermined, then the user should enter **0** (zero).

#### ENTER THE NUMBER OF STRATA FOR THIS RUN?

The program allows for a maximum of 12 strata. If the estimates are for an unrestricted sample, then the user needs to enter a 1.

---



---

	<b>ESTIMATED</b>	<b>EST</b>
<b>STRATUM</b>	<b>STRATUM NAME</b>	<b>ESTIMATED MEAN UNIVERSE SIZE</b>
		<b>STDRD DEV.</b>

For each stratum, the user will be prompted to first enter a brief description of the stratum. A maximum of 20 characters will be printed for each stratum. This entry is optional and the user may just hit the Enter key instead of entering a description.

The user needs to enter 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 to the user in estimating the mean.

The universe size of the stratum is usually known by the user. If, however, the universe is an unknown, then the best estimate of the universe size needs to be used.

The estimated standard deviation may be the hardest for the user to calculate. The standard deviation is a measurement of the variance of the items about the mean. By adding and subtracting the standard deviation from the mean, an interval is calculated. Approximately 68 percent of the items sampled should fall within this range if the universe is normally distributed.

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

The better the estimates of mean, universe and standard deviation, the more accurate the estimates of sample sizes will be.

### **DO YOU NEED TO EDIT YOUR ENTRIES (Y/N)?**

The user may edit the entries after all the strata values have initially been entered. The query requires a **Y** or **N** response.

### **ENTER THE STRATUM NUMBER OR 00 TO QUIT?**

If the user responds with a **Y** to the prior query, then the user will be asked to enter a stratum number for values to be edited. An entry of two zeroes (00) will exit the user from the modification routine.

### **<===ENTER NAME FOR STRATUM**

The user may rename the stratum. If the user merely hits the Enter key, then the stratum name will remain unchanged.

---

---

### <===ENTER MEAN FOR STRATUM

The user may now edit the mean value. If the user merely hits the Enter key, then the mean value will remain unchanged. The user may include commas in the mean value.

### <=ENTER UNIVERSE

The user may now edit the universe value. If the user merely hits the Enter key, then the universe value will remain unchanged. The user may include commas in the universe size.

### CHANGE STANDARD DEVIATION

While the above prompt does not appear, the cursor will now blink below the existing standard deviation amount. The user may change the existing amount or just hit the Enter key to leave the standard deviation unchanged. The user may include commas in entering the standard deviation amount.

An example of the query screen appears on page 5-10. The user responses are shown in boldface. An example of the output appears on page 5-10.

### Program Output

The program will repeat the entries made by the user and then generate a series of sample sizes based on the entries. The estimates will be given at both the 90 and 95 percent confidence level for precisions from 5 to 50 percent. The precision will be at 5 percent increments. As soon as the overall sample size at the 90 percent confidence level falls below 30, then the program will stop generating sample sizes.

When the sample size for any given stratum falls below 30, then a message "**SEE COMMENTS BELOW**" will appear. The note at the bottom of the output states:

#### NOTE:

THE SAMPLE SIZES GENERATED ABOVE WERE THE RESULT OF MATHEMATICAL FORMULAS AND DID NOT INCORPORATE MANAGEMENT DECISIONS CONCERNING THE PURPOSE OF THE SAMPLE OR CURRENT SAMPLING POLICIES OF OUR ORGANIZATION. THEREFORE YOU MAY NEED TO INCREASE THE SAMPLE SIZES IN ORDER TO BE IN COMPLIANCE WITH MANAGEMENT'S OBJECTIVES.

If the programs calculates a sample size that is greater than the universe, then a message "**SEE COMMENTS BELOW**" will appear. The note at the bottom of the output states:

NOTE:

THE FORMULAS CALCULATED A SAMPLE SIZE GREATER THAN THE UNIVERSE. 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 the user entered a total sample size to be allocated, the formulas may calculate something slightly different in total due to rounding. If this situation occurs, then a message "**SEE COMMENTS BELOW**" will appear. The note at the bottom of the output states:

NOTE:

DUE TO ROUNDING, THE TOTAL SAMPLE SIZE CALCULATED (#,###) DOES NOT EQUAL THE SAMPLE SIZE REQUESTED (#,###)

The output also displays the percentage of the sample that appears in any given stratum. These percentage could be applied by the user to the overall sample size that is finally selected. Based on the entries made by the user, these ratios represent the optimum allocation of the sample.

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 07/27/92          SAMPLE SIZE ESTIMATORS          TIME:09:20

ENTER THE NAME OF THIS AUDIT/REVIEW?  SAMPLE SIZE TEST

ENTER TOTAL SAMPLE SIZE, ELSE 0 (ZERO)  200

DO YOU WANT RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?  P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?

ENTER THE NUMBER OF STRATA FOR THIS RUN?  2

STRATUM  STRATUM NAME      ESTIMATED MEAN  ESTIMATED
          HIGH INCOME      10,000.00      UNIVERSE SIZE  ESTIMATED
          LOW INCOME       5,000.00      100,000        STANDARD DEVIATION
          5,000.00      500,000        5,000.00
          4,000.00

DO YOU NEED TO EDIT YOUR ENTRIES (Y/N)?  N
    
```

```

DEPARTMENT OF HEALTH & HUMAN SERVICES
OIG - OFFICE OF AUDIT SERVICES
DATE: 07/27/92          SAMPLE SIZE ESTIMATORS          TIME: 09:20

AUDIT/REVIEW: SAMPLE SIZE TEST

THE ESTIMATORS ARE BASED ON THE FOLLOWING ENTRIES:

NBR      DESCRIPTION      --MEAN--      ---UNIVERSE---      STANDARD DEVIATION
  1      HIGH INCOME      10,000.00      100,000              5,000.00
  2      LOW INCOME       5,000.00      500,000              4,000.00
=====

STRATUM      RATIO      SAMPLE
              SIZE
HIGH INCOME      20.00%      40
LOW INCOME       80.00%      160

TOTAL SAMPLE SIZE REQUESTED      200

PRECISION AT 90% CONFIDENCE LEVEL      +/-      8.31%
PRECISION AT 95% CONFIDENCE LEVEL      +/-      9.90%
    
```

## ADD DECIMAL POINTS

### Purpose

This program allows users to insert decimal points into files downloaded from mainframe files. Normally files accessed on mainframe systems with application languages such as COBOL do not store the decimal point in numeric fields. When these files are downloaded into database packages on microcomputers, the decimal points must be added by the user through a program on either the mainframe or microcomputer or by re-entering selected fields in the microcomputer database.

This program assumes that the location of decimal points will be consistent throughout the file. The utility should be used after the file has been downloaded and before it is placed into a database package. A maximum of 20 decimal points may be added with this utility to each record.

### Input Queries

#### ENTER NAME OF FILE NEEDING DECIMAL POINTS?

The user must enter the name of the file downloaded from the mainframe. If the file is not in the current directory, then a complete path to the data file must be given. For example, if the data file was on a floppy disk in the "A" drive, the user could enter **A:\MAINFRME.DAT**.

If the data file cannot be located, the user will be prompted to re-enter the path and file name.

#### ENTER NAME OF NEW FILE WITH DECIMAL POINTS?

The user needs to assign a name to the new file that will be created with this program. The file name may be up to eight characters and have an extension by adding a decimal point and up to three characters. The file must not have the same path and file name as the input file.

After this response is entered, the program will clear the screen and then bring the first record from the downloaded file to the screen. Using this record as a reference, the user will then be prompted.

**FOR EACH DECIMAL POINT TO BE ADDED, YOU WILL BE PROMPTED FOR THE POSITION IN THE CURRENT RECORD THE DECIMAL POINT SHOULD FOLLOW. WHEN FINISHED JUST HIT THE <ENTER> KEY AT THE NEXT PROMPT. DECIMAL POINT LOCATION NUMBER 1**

As the user identifies each position in the record that is to be followed with a decimal point, the position of the first record displayed in the top half of the screen will blink and the number will appear in the lower portion of the screen. If the position entered by the user is not numeric or greater than the length of the record, the user will be prompted for the next entry. If the position has previously been entered by the user, the program will just prompt for the next location.

**TO MAKE ANY CHANGES TO THE LOCATION OF THE DECIMAL POINTS ENTER THE TYPE (<A>DD, <D>ELETE, <Q>UIT) AND THE RECORD POSITION (e.g. A123)**

The user now has the option to add or delete locations previously entered. As each modification is made, the change will be reflected on the record displayed by either turning on or off the blinking at the appropriate location.

When the user enters "Q", then the points will be added to the first record and the record will be displayed as it will look in the new data file.

**DO YOU WANT TO CONTINUE PROCESSING ALL RECORDS (Y/N)?**

After reviewing the new record, the user has the option to process all of the records in the same manner or not to process the file.

**DO YOU WANT TO EXIT THIS PROGRAM (Y/N)?**

If the user did not want to continue processing the file, then the user has the option of re-running this program or going back to the host program ("SAMPLE").

An example of the input screens is shown on page 5-14. The user entries are displayed in **boldface**. An example of the screen after the records have been processed is shown on page 5-15.

## Program Output

As the file is being processed, the program will display the current number of the record being processed and the current number of the record being written to the new data file. When finished processing, the program will display:

(PRESS ANY KEY TO CONTINUE)

The user should verify that the number of records displayed in the counts on the screen reconcile to the totals downloaded from the mainframe file.

```

                                DEPARTMENT OF HEALTH & HUMAN SERVICES
                                OIG - OFFICE OF AUDIT SERVICES
DATE: 03/18/93                ADD DECIMAL POINTS TO A RECORD                TIME:17:08
                                VERSION: MONTH YEAR

ENTER NAME OF FILE NEEDING DECIMAL POINTS? B:TESTFILE.DWN
ENTER NAME OF NEW FILE WITH DECIMAL POINTS? B:TESTFILE.PNT

=====NEW SCREEN=====

                                DEPARTMENT OF HEALTH & HUMAN SERVICES
                                OIG - OFFICE OF AUDIT SERVICES
DATE: 03/18/93                ADD DECIMAL POINTS TO A RECORD                TIME:17:08
                                VERSION: JANUARY 1993
                                INPUT FILE: B:TESTFILE.DWN

.....5....10...15...20...25...30...35...40...45...50
 1          50
001147669Q31705735299005404638950002609/01/21MU01/
51          99    30/8701/30/8701/30/87000000000000000000236002/26/87

FOR EACH DECIMAL POINT TO BE ADDED, YOU WILL BE PROMPTED FOR THE POSITION
IN THE CURRENT RECORD THE DECIMAL POINT SHOULD FOLLOW. WHEN FINISHED JUST HIT THE <ENTER> KEY
AT THE NEXT PROMPT.

DECIMAL POINT LOCATION NUMBER 2? 89

***** LOCATIONS ALREADY ASSIGNED *****
79

TO MAKE ANY CHANGES TO THE LOCATION OF THE DECIMAL POINTS ENTER THE TYPE
(<A>DD,<D>ELETE,<Q>UIT) AND THE RECORD POSITION (E.G. A123) Q

DO YOU WANT TO CONTINUE PROCESSING ALL RECORDS (Y/N)? Y

```

```

                                DEPARTMENT OF HEALTH & HUMAN SERVICES
                                OIG - OFFICE OF AUDIT SERVICES
DATE: 03/18/93                ADD DECIMAL POINTS TO A RECORD                TIME: 17:08
                                VERSION: MONTH YEAR
                                OUTPUT FILE: B:\TESTFILE.PNT

. . . . 5 . . . 10 . . . 15 . . . 20 . . . 25 . . . 30 . . . 35 . . . 40 . . . 45 . . . 50
  1                               50
001147669Q31705735299005404638950002609/01/21MU01/
  51                               100
30/8701/30/8701/30/8700000000.0000000023.6002/26/8
101                               101    7

                                B:\TESTFILE.DW
N
                                B:\TESTFILE.PNT
                                RECORDS
READ:                          403
                                RECORDS WRITTEN: 403

***** LOCATIONS ALREADY ASSIGNED *****
79      89
(PRESS ANY KEY TO CONTINUE)

```

## TRANSPOSITION ERROR IDENTIFICATION

### Purpose

This program assists the user in locating transposition errors. Such errors occur when two digits in a number are reversed in the process of being transcribed. For example, 85 is recorded as 58. This type of error is identified when the total of a transcribed series of numbers differs from the original total by an amount evenly divisible by nine. In the above example, the difference between the two numbers is 27 (85 - 58). This difference is evenly divisible by nine ( $27 / 9 = 3$ ). However, not all differences divisible by nine are the result of transposition errors. This program will identify the pair(s) of digits that could have been transposed and the location of the digits within a number. While transposition errors are commonly the result of adjoining digits being switched, this program will also identify non-adjoining digit switching.

This program is based on an article in the June 1992 issue of The Practical Accountant that was written by Jay B. Abrams.

### Input Queries

#### ENTER THE AMOUNT OF THE DIFFERENCE?

The user is asked to enter the amount of the difference between the total of an original series of numbers and the total of the transcribed series of numbers. The difference may be entered as a negative or a positive amount. The entry may contain dollar signs, commas or other non-numeric symbols.

The program will test to determine if the difference amount is divisible by nine. If the difference is not divisible by nine, an error message will be printed.

#### DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN?

The user selects the appropriate output by responding to the query with a **D**(disk), **P**(printer) or **S**(screen).

#### ENTER NAME OF FILE TO RECORD THE RESULTS?

If the user selects the disk for the output, the program will prompt for a file name. The primary file name may be up to eight characters. The user may add an extension with a period and a maximum of three characters. If the disk file is to be written to another drive and/or directory, the user must include the path in the file name. For example, a

disk file going to the "A" drive could be **"A:FILEONE.OUT."**

By sending the output to disk, the user may then transmit the information by telecommunications to another microcomputer or store the output when a printer is not available. To print the disk file, the user could enter **TYPE FILEONE.OUT** to have the results displayed on the screen or **TYPE FILEONE.OUT > PRN** to have the results sent to the printer. The user could also retrieve the file through a word processing package such as WordPerfect.

### **ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2)?**

If the user selects the printer for output, then this query will appear. Computers that use multiple parallel ports or have connection through a LAN to multiple ports may route the output to the appropriate port. The user only needs to enter the port number between 1 and 4. If the output is going to LPT1, then the user has the option of only hitting the Enter key.

An example of the input screen is shown on page 5-18. The user entries are in **boldface**. An example of the requested output is also shown on page 5-18.

### **Program Output**

The program will generate two columns of numbers showing the potential transposition errors. The first column displays debit errors while the second column displays credit errors. In determining the amount of the difference initially, the user should subtract the transcribed total from the original total. In the example above, the difference was 27. The user would then search through the original series of numbers looking for the seven possible transpositions. One of the five potential transpositions displayed under the debit column is 85. For each match in the original column with a matching pattern in the debit column, the user should look in the transcribed column to see if the digits have been transposed.

However, if the difference was -27, then the credit column would be used to identify the patterns to be tested in the original column of numbers.

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
TRANSPOSITION ERROR IDENTIFICATION

DATE: 09/21/92 TIME: 17:13

ENTER THE AMOUNT OF THE DIFFERENCE 27

DO YOU WANT THE RESULTS PRINTED TO <D>ISK, <P>RINTER OR <S>CREEN? P

ENTER PRINTER PORT NUMBER, IF OTHER THAN LPT1 (E.G. 2) ?

DEPARTMENT OF HEALTH & HUMAN SERVICES  
OIG - OFFICE OF AUDIT SERVICES  
TRANSPOSITION ERROR IDENTIFICATION

DATE: 09/21/92 TIME: 17:13

DIFFERENCE AMOUNT = 27

DEBIT ERRORS	CREDIT ERRORS
x,xxx,x96	x,xxx,x69
x,xxx,x85	x,xxx,x58
x,xxx,x74	x,xxx,x47
x,xxx,x63	x,xxx,x36
x,xxx,x52	x,xxx,x25
x,xxx,x41	x,xxx,x14
x,xxx,x30	x,xxx,x03

YOUR TRANSPOSITION ERROR OCCURRED IN THE COLUMNS WHERE THE DIGITS ARE PRINTED. CHECK FOR ONLY THOSE DIGIT COMBINATIONS THAT ARE PRINTED ABOVE.