RAT-STATS 17

Statistical Software

Overview

A Challenge Solution Submitted by

J&J Cyber Security LLC

in response to

Dept. of Health and Human Service's

"Statistical Software for Healthcare Oversight" Challenge

Prepared By:

J&J Cyber Security LLC
POBOX 9585
Moscow ID 83843
(208) 883-8181

Executive Summary

This document discusses the Challenge Solution submitted in response to the Dept. of Health and Human Services' "Statistical Software for Healthcare Oversight" challenge. As stated in the Challenge Description.

> *The objective of the current challenge is to develop the foundation for an upgraded version of RAT-STATS. The current version of RAT-STATS is well validated; however, its user interface can be difficult to navigate and the underlying code makes the software costly to update. OIG needs a new, modern version of the software that is easy to use and can be extended in a cost effective manner. In addition, the new version of the software must be 508 compliant.*

This submission provides the required 4 functions specified in the challenge.

This submission also provides a Graphical User Interface (GUI) like that of the current RAT-STATS program.

This submission also supports input and output from Excel .xlsx files. This support is provided through use of library routines and does not actually run Excel. This submission does not support input or output to MS Access database. A user should be able to import the Excel files into Access.

This submission also does not provide an updated user manual or online help.

This submission contains the following files:

(1)	JJCyber-RATSTATS.exe.zip	A zip file containing the executable solution and related DLL files. This solution replicates the 4 target RAT-STATS functions: Single Stage of Random Numbers, Unrestricted Attribute Appraisal, Unrestricted Variable Appraisal, and the Stratified Variable Appraisal.
(2)	JJCyber-RATSTATS.src.zip	A zip file containing the source code for the executable, header files, resource files, makefile and relevant DLLs.
(3)	JJCyber-Overview.docx ***This Document***	A document providing a detailed discussion of the submission, including a discussion of differences between the original RAT-STATS and this submission.
(4)	JJCyber-Devel.docx	A document describing the overall code structure and design with guidance how to modify the solution to add additional functionality.
(5)	JJCyber-Licenses.docx	A document listing all software licenses associated with the source code and libraries used as part of the project.
(6)	JJCyber-Summary.docx	A document providing summary submission information as requested on the challenge website.
	README.txt	A simple text version of the preceding information.

Contents

List of Figures

1 Project Overview

1.1 Background

Work began on this submission at the end of November 2016. The first part was to experiment with translating the Visual Basic and R software into different languages to determine ease of translation, functional equivalence and performance. Tests with Python and Java led to the conclusion that these interpreted languages had too many performance issues for confidence interval calculations. Tests with embedding native 'C' code with Python produce acceptable performance results, but greatly increased the maintainability of the code.

A final decision was made to implement the code in C++ (with limited use of objects). This also enabled the use of standard Microsoft tools and controls. The solution is implemented using freely available tools to simplify future maintenance.

The submission is not commercial release/production quality. However, it has been designed to allow additional testing and modification to allow it to grow to production quality.

1.2 Features

This solution to the challenge was designed to utilize the standard Windows Desktop interface and user customizable attributes. As can be seen in the screen shots, it utilizes the standard windows desktop display, with the standard minimize, maximize and exit icons. It utilizes the standard menu bars, both mouse and keyboard selectable.

1.3 Risks / Future Work

This submitted solution does not provide the full functionality of the current RAT-STATS application, and requires a substantial of work to include all of these functions and to become a fully functional product. The following summarize main issues with respect to this submission and future expansion.

English Only: This application has not been internationalized, and was not designed to support multiple languages. Therefore, there is no support for localization of character sets.

Development: The full solution utilizes several technologies, which may complicate future development and support. The implementation is written in C++, with a strong leaning on standard C coding styles. The GUI interface utilizes standard MSDN Windows controls and dialogs. The expansion of these is documented, however the implementation does not tie into a specific GUI development environment. Therefore the placement of the controls is a manual process. This is documented with several examples, such that an experienced C/C++ programmer should be able to expand the software. Place-holders for unimplemented menu items have been placed in the code and GUI. *JJCYBER-Devel* guide provides detailed instructions for expansion.

User Testing: A well-designed program should undergo extensive user testing. This solution has not gone through that process. The design of the GUI interface should allow for relatively simple modifications to improve the user experience. The solution provides a couple of different GUI options, which will be discussed further, that should be evaluated during user testing.

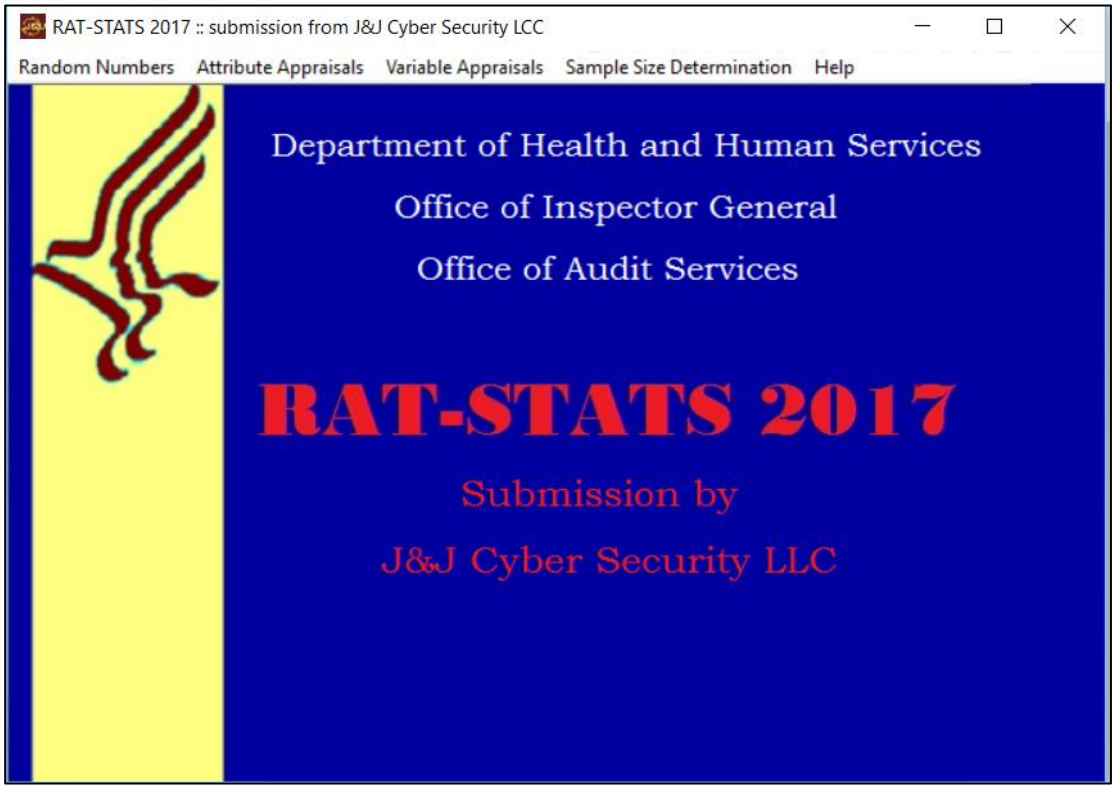

Figure 1: RAT-STATS 17 Main Screen

2 Target RAT-STATS Functions

This section discusses the target RAT-STATS functions implemented in this project as specified in the Challenge rules: Single Stage of Random Numbers, Unrestricted Attribute Appraisal, Unrestricted Variable Appraisal, and the Stratified Variable Appraisal.

Each of these functions can be selection from drop-down menus from the main screen (Figure 1). Menu placement is designed to mirror placement on the current version of RAT-STATS. The drop-down menus currently include the requested functions as well as disabled place-holders for additional functionality.

To address accessibility issues, menus can be selected with a mouse, or by hitting the <Alt> key. This is standard Microsoft practice. The <Alt> key will allow subsequent <TAB>ing through the menu options and then hitting <Enter> or <↓> (down arrow). In addition, a menu can be selected using a leading character (such as "V" for Variable Appraisal) to select the menu. The characters to be used are shown underlined after the <Alt> key is selected.

2.1 Random Numbers

The Challenge Requirements request implementation of the *Single Stage of Random Numbers* functionality. To demonstrate expansion and flexibility, this submission also implements sets of 2, 3 and 4 random numbers. The implementation, as described in *JJCyber-Devel* document, is parameterized for the number of variables, allowing for code reuse. The screen layouts are consistent between the different options (Figure 2 and Figure 3).

Figure 2 shows some of the user interface options that are available.

- *Static Data Types:* First is the grayed text in the data-input boxes. This text is intended to give guidance to the type of data required in the box. Although this is useful, it may be distracting to end users. Therefore, it is only implemented on this one screen to demonstrate the capability, and to help with user testing.
- *Pop-up guidance*: In addition to the static typing, we have the option of a pop-up box (the rectangle "Enter optional seed") that provides guidance for users.
- *Warnings:* If a user enters data that is not of the correct data type (such as the "45a") entered for # Generated Randomly, a Warning Icon appears next to the box. In addition, a pop-up balloon will appear with instructions for the user. This includes a warning the Hi frame value is not greater than the low frame, or the frame size is too small for the desired quantity of random numbers (Figure 3).
- *Selectively Enabled Save*: The Save As.. button is disabled until all mandatory fields are filled in, and data type checks are correct.

Once the correct values are selected, the "Save As.." button is enabled. The user can select an output file. If the filename ends in ".xlsx", it is assumed that the user wants to save as an Excel File, and the data is saved in that format, consistent with the format generated by the current RAT-STATS program. Any other file extension will result in saving the file as a plain text file in the format layout used by the current RAT-STATS solution. This will simplify compatibility with other applications that require specific spacing in that file.

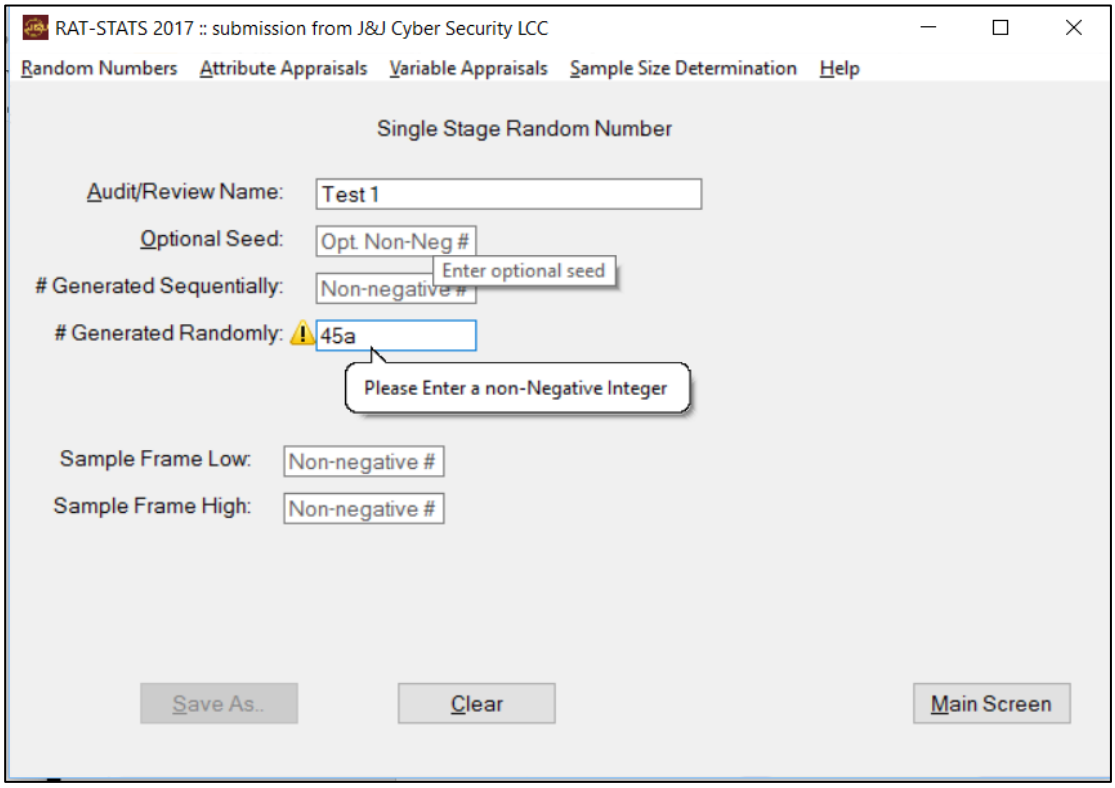

Figure 2: Single Stage Random Number Input Screen

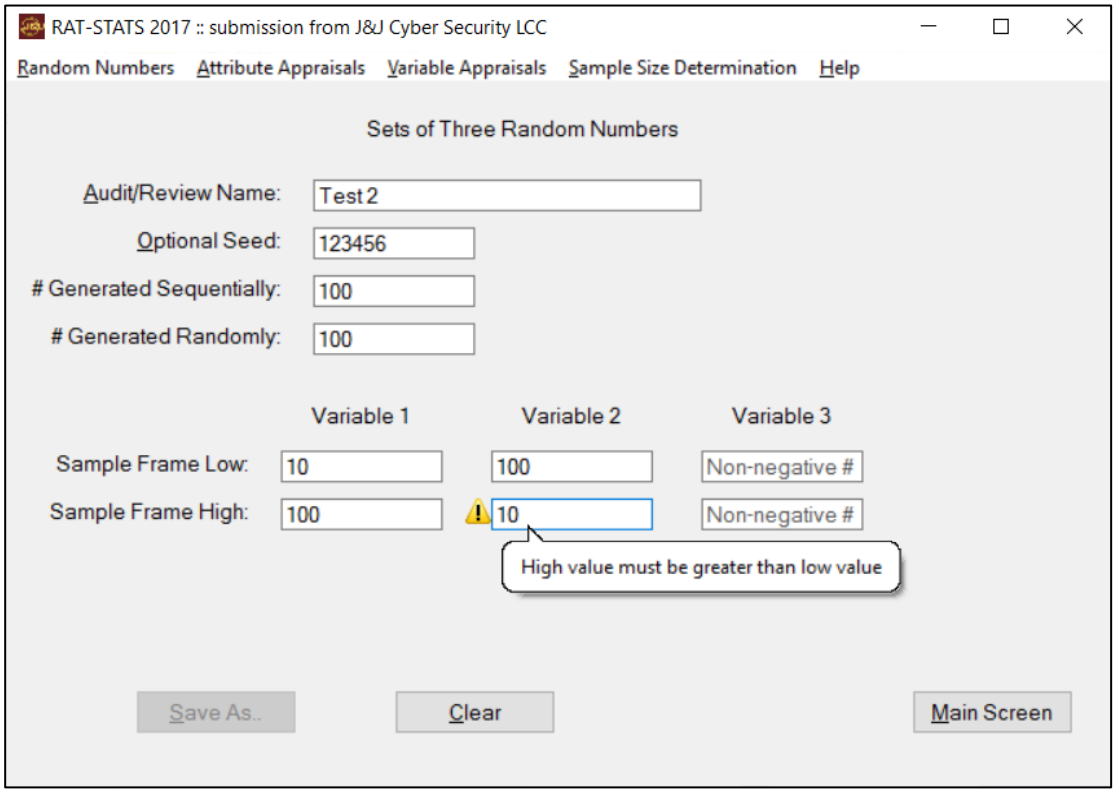

Figure 3: Sets of Three Random Number Input Screen

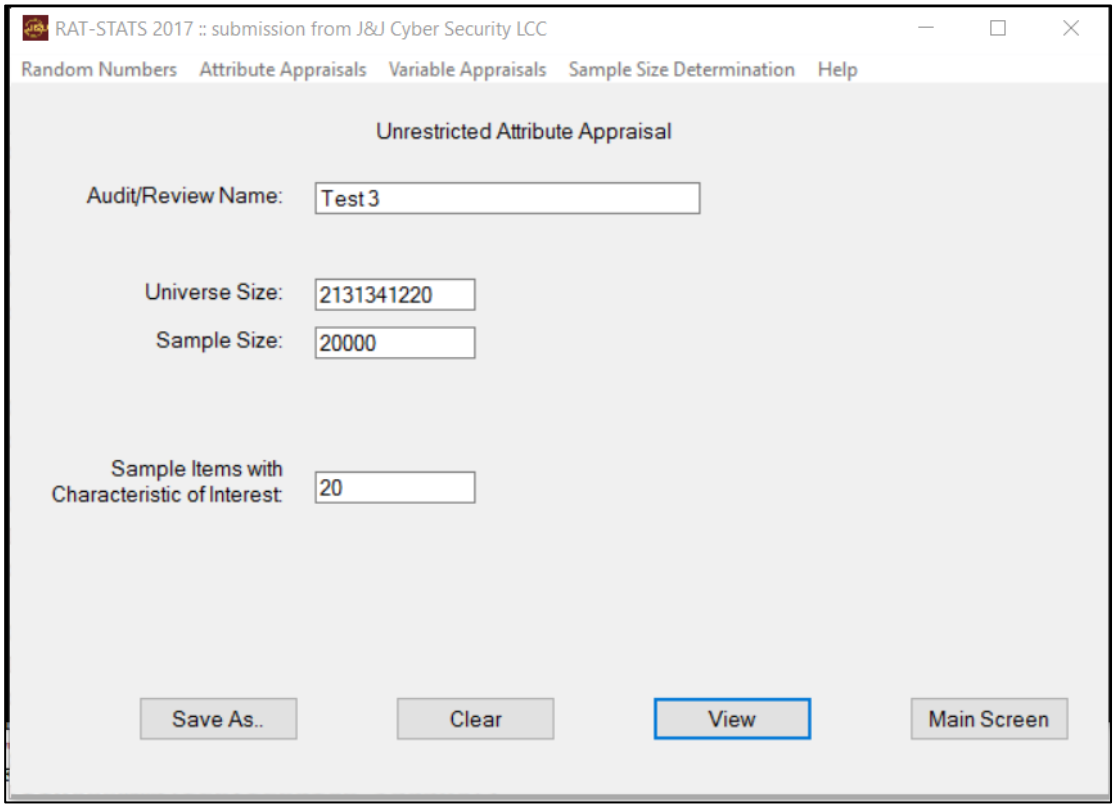

Figure 4: Unrestricted Attribute Appraisal Input Screen

2.2 Unrestricted Attribute Appraisal

The Unrestricted Attribute Function (Figure 4 and Figure 5) is implemented similar to the oringal RATS-STATS implementation. Users input the audit/Review name, universe size, sample size and # of sample items with the characteristic of interest. Then they can view the results (Figure 5) on the screen or printed to a selected text file.

There is no option to output to an Excel file.

There is no option to view a one-sided confidence interval.

There is no option to print results for a single confidence level. This should be easy to add to the codebase.

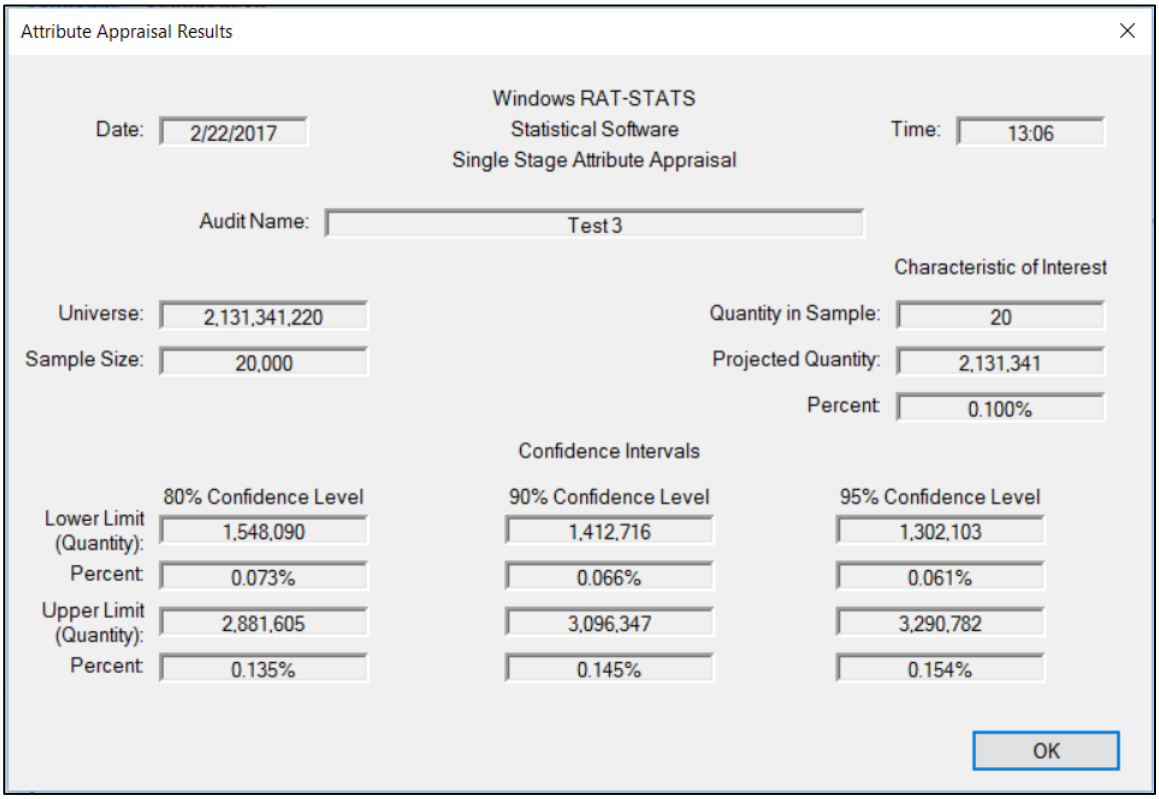

Figure 5: Unrestricted Attribute Appraisal Output Screen

2.3 Unrestricted Variable Appraisal

The Variable Analysis functions have some similar functionality, and therefore similar user interface screens.

The first input screen (Figure 6) allows the user to select the input file, specify parameters and displays summary results. The sample size will be automatically calculated from the input file. The summary results will also be calculated from the input file.

When the users clicks the "Select Input File" button, they are provided with the standard Windows Open File dialog, with filtering for ".xlsx", ".txt" and all files. If the user selects a MS Excel input file, they are given the opportunity to select starting Excel Spreadsheet, rows and columns of the input data (Figure 7), otherwise they will select starting row and column of the data from a text file (Figure 8).

The Excel input consists of 3 check boxes. The user checks the box for a type of data, then clicks on the data entry box (where it says "Click to Choose" to enable the Excel information boxes at the bottom of the screen. These boxes summarize the WorkSheets in the file along with the top few rows of data.

The user can select the Workbook Sheet and the starting row and column of the desired input data. The resulting Workbook Sheet name and Excel Cell identifier (e.g, A1, C2..) will appear in the boxes. When the user is finished, they can click on the "Done" button, which returns them to the data input screen with the summary results and sample size boxes filled in. This input format differs from the Challenge Specification that requests a specific "Flag" to indicate if the input data contains header

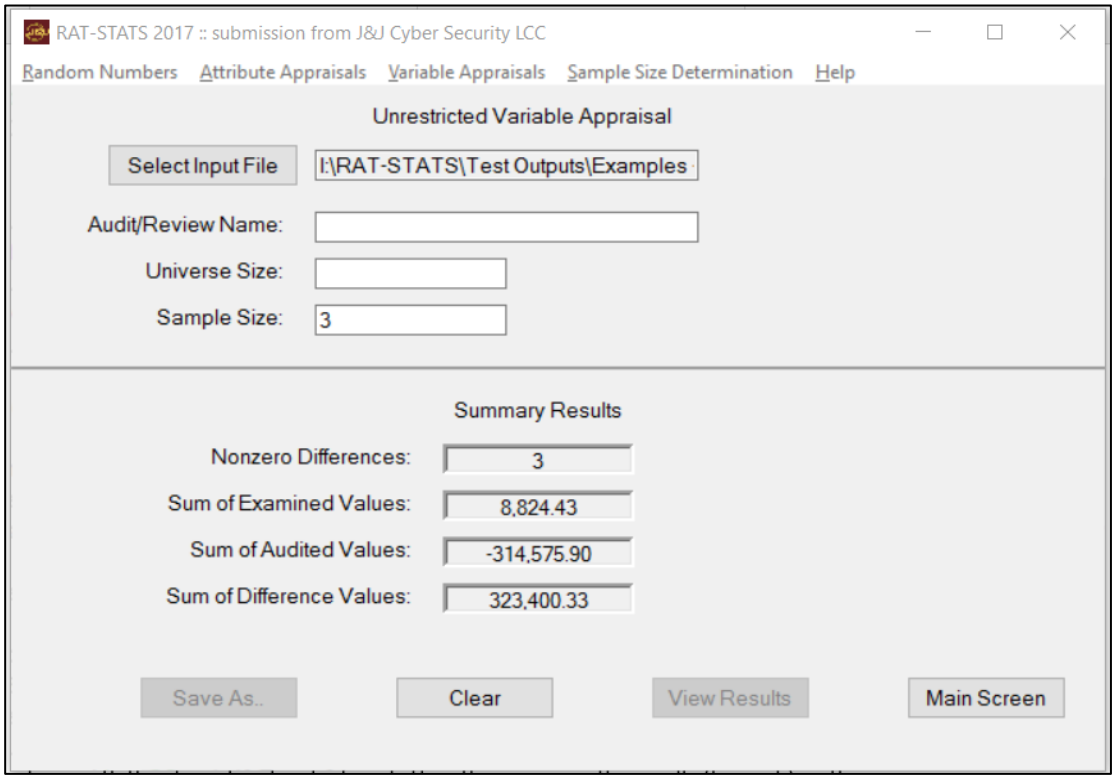

Figure 6: Unrestricted Variable Appraisal Initial Input & Summary Screen

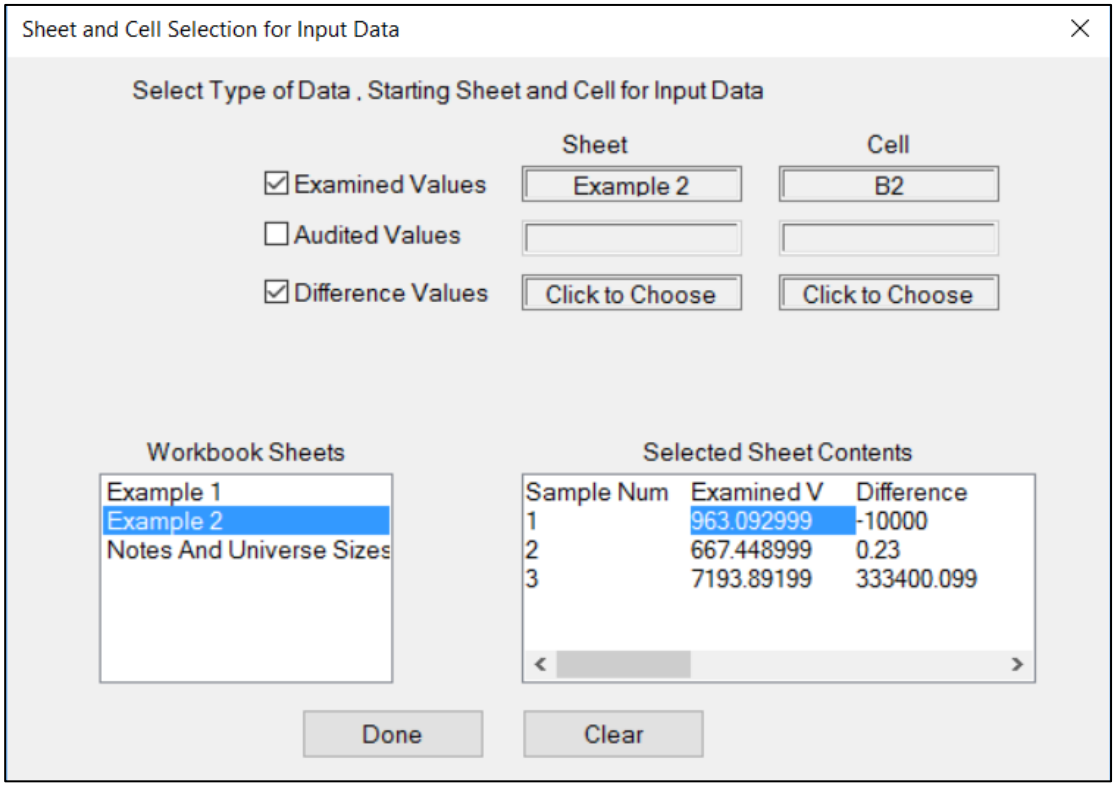

Figure 7: Excel Data Selection for Unrestricted Variable Appraisal

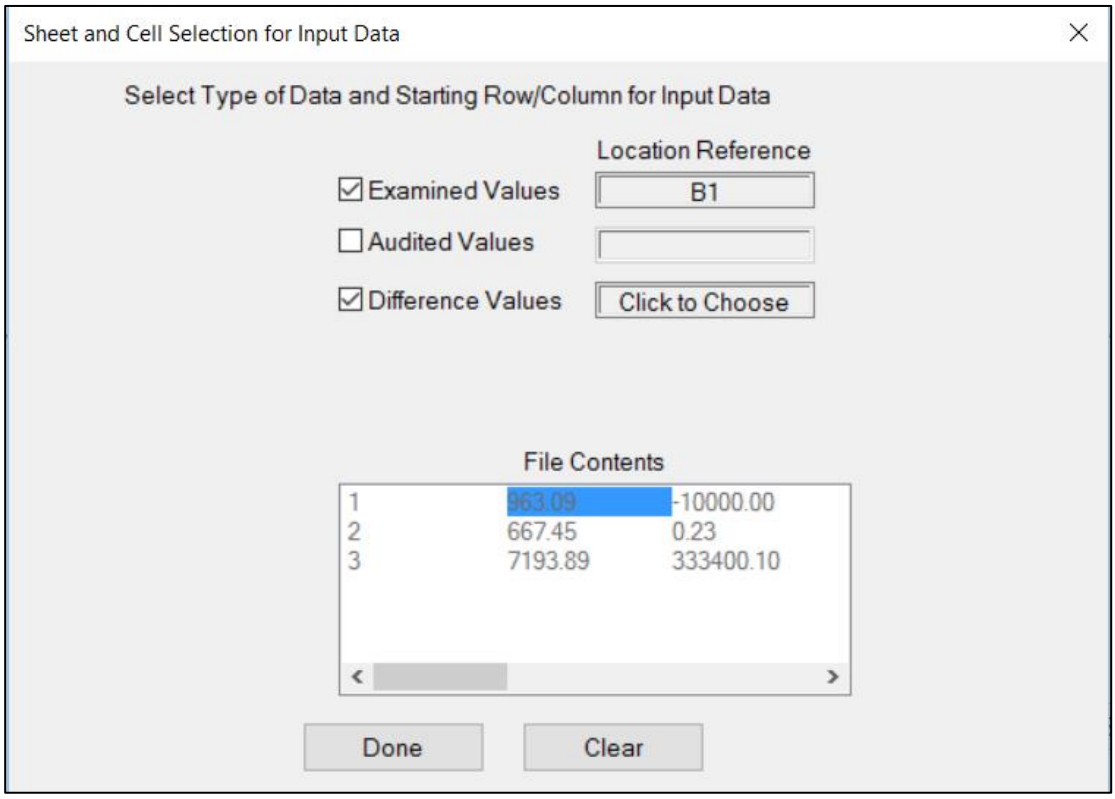

Figure 8: Text File Data Selection for Unrestricted Variable Appraisal

information. It is compatible with existing input file formats for the current RAT-STATS software, but allows for additional future flexibility, such as not requiring data to start on only first or second row/column, and not requiring examined data to be in the column preceding Audit or Difference values.

If the input file is not an Excel File, the user selects the appropriate starting rows and columns from the text file, similar to the Excel selection input. The text input file must be space/tab separated. We have expanded the acceptable format options to allow numbers with thousand separators (comma's) for a data entry.

Once the user enters the correct Universe size and Audit name, they can select to View the results on screen (Figure 9), or save them to a text file, or Excel file.

There is a small difference between the outputs shown here and the current version of RAT-STATS. The results in the example shown in Figure 9 require 64 bit integers to allow for these large limits. RAT-STATS will usually show a '-' as the value for numbers that are out of range.

If the user clicks on the "Additional Information" button, they will scroll through the results for Examined, Audited and Difference values.

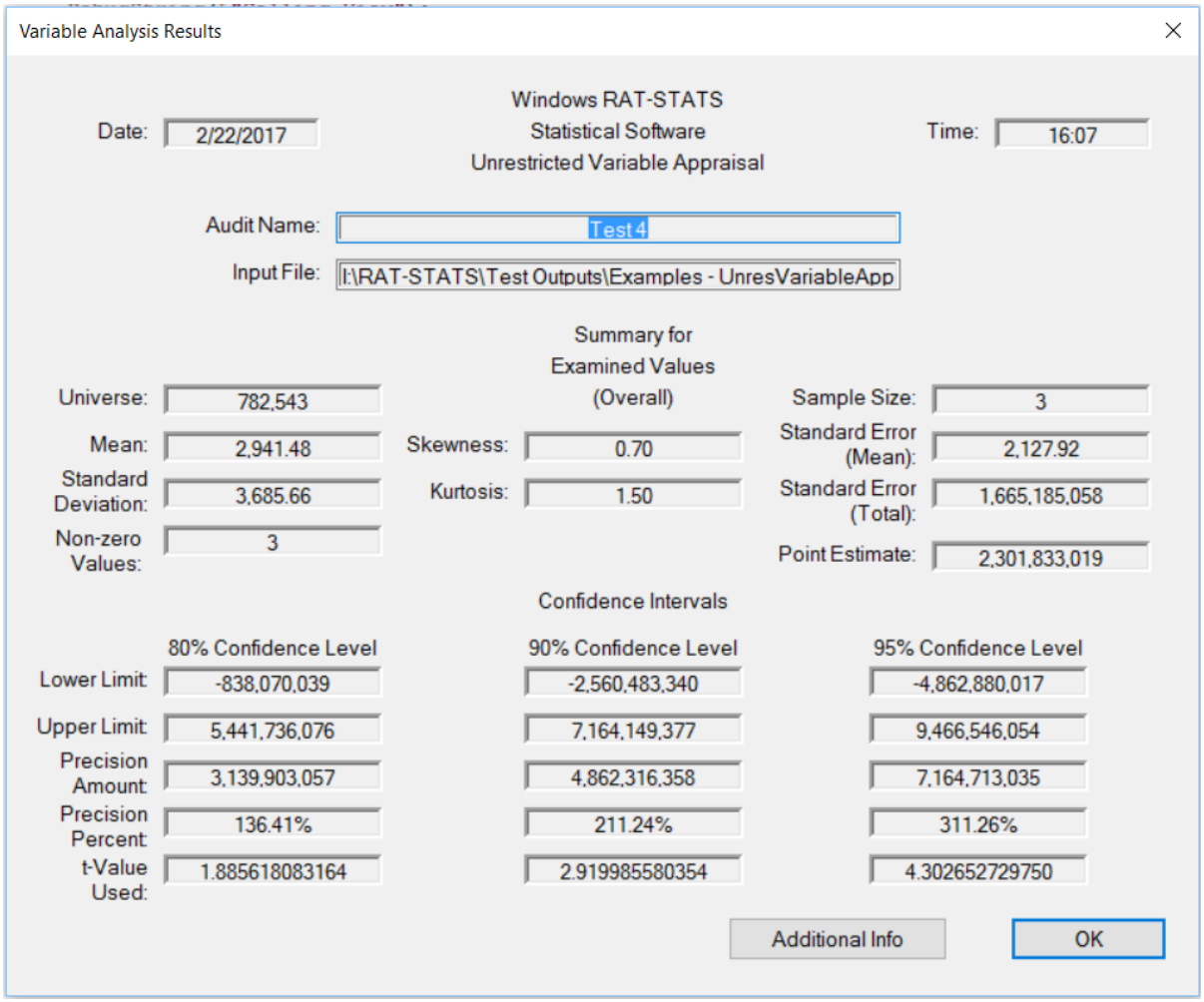

Figure 9: Unrestricted Variable Appraisal Results Screen

2.4 Stratified Variable Appraisal

The first input screen (Figure 10) allows the user to select the input file, specify parameters and displays summary results. The sample size, universe size and number of strata will be automatically calculated from the input file. The overall summary results will also be calculated from the input file.

When the users clicks the “Select Input File” button, they are provided with the standard Windows Open File dialog, with filtering for “.xlsx”, “.txt” and all files. If the user selects a MS Excel input file, they are given the opportunity to select starting Excel Spreadsheet, rows and columns of the input data, as wellas thte sheet, rows and columns of the strata data (Figure 11). If the user selects a text file, they will see a similar screen to select row and column of the data, but will also have to select a separate input file for the strata sizes data entry (Figure 12).

The Excel input consists of 5 check boxes, 3 for data and 2 for strata sizes. The user checks the box for a type of data, then clicks on the data entry box (where it says “Click to Choose” to enable the Excel information boxes at the bottom of the screen. These boxes summarize the WorkSheets in the file along with the top few rows of data.

The user can select the Workbook Sheet and the starting row and column of the desired input data. The resulting Workbook Sheet name and Excel Cell identifier (e.g, A1, C2..) will appear in the boxes. When the user is finished, they can click on the "Done" button, which returns them to the data input screen with the summary results and sample size boxes filled in. This input format differs from the Challenge Specification that requests a specific "Flag" to indicate if the input data contains header information. It is compatible with existing input file formats for the current RAT-STATS software, but allows for additional future flexibility, such as not requiring data to start on only first or second row/column, and not requiring examined data to be in the column preceding Audit or Difference values.

If the input file is not an Excel File, the user selects the appropriate starting rows and columns from the text file, similar to the Excel selection input. They must also click on the buttonto select the separate strata sizes file. The text input files must be space/tab separated. We have expanded the acceptable format options to allow numbers with thousand separators (comma's) for a data entry, to permit a different order of columns, and we do not require the first column to have identifiers or row numbers. Once the user enters the correct Audit name, they can select to View the results on screen (Figure 13), or save them to a text file, or Excel file.

If the user clicks on the "Additional Information" button, they will scroll through the results for Examined, Audited and Difference values. If the user selects "Next Stratum" or "Prev Stratum" buttons they will cycle through the various stratum, and can select additional information for each. The "Overall" button shows the combined statistics of all strata.

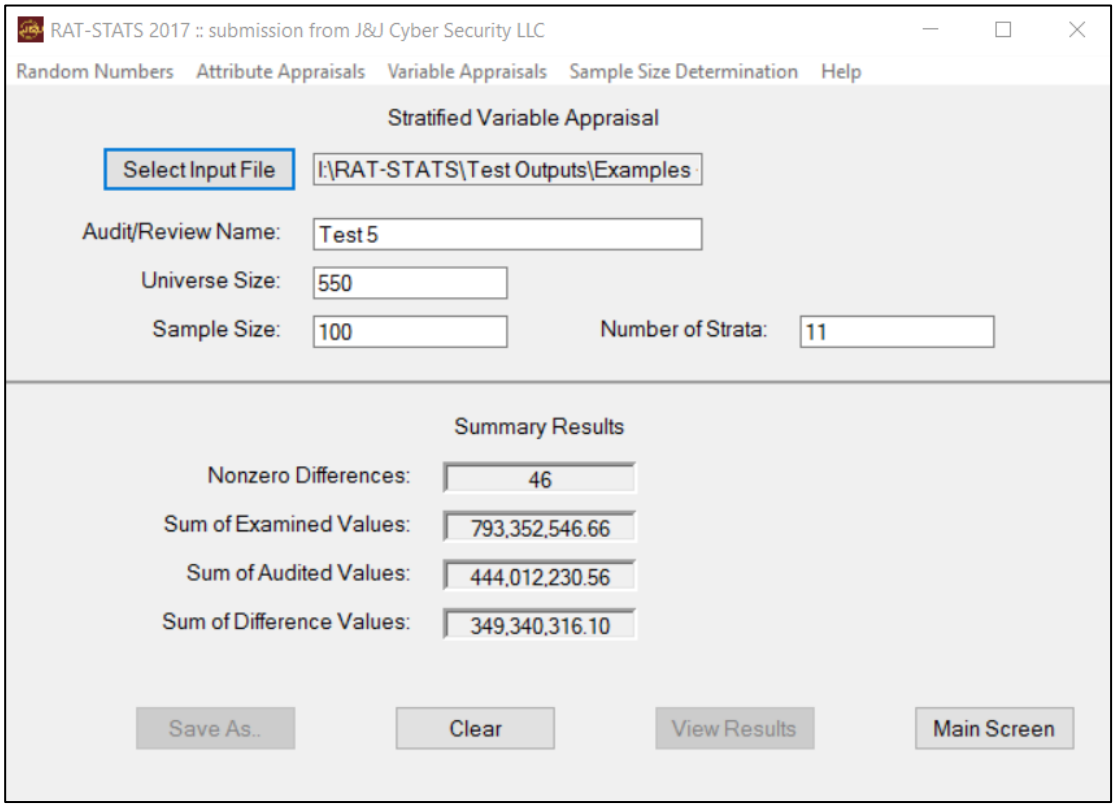

Figure 10: Stratified Variable Appraisal Initial Input & Summary Screen

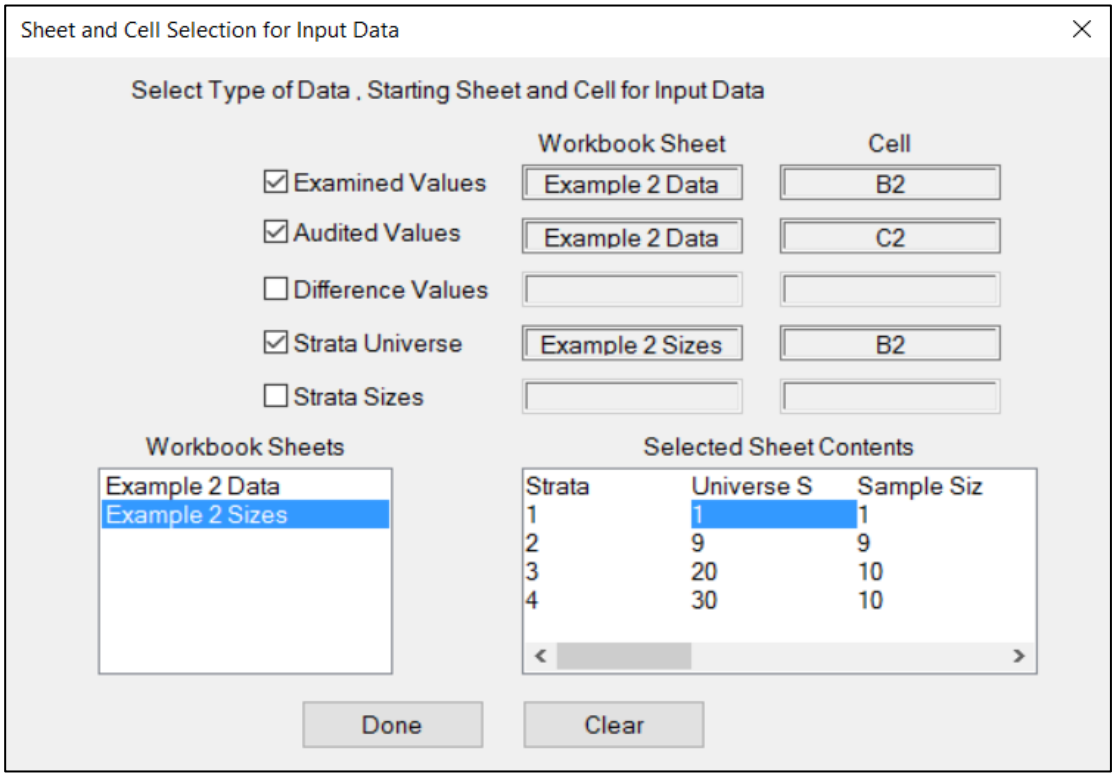

Figure 11: Excel Data Selection for Stratified Variable Appraisal

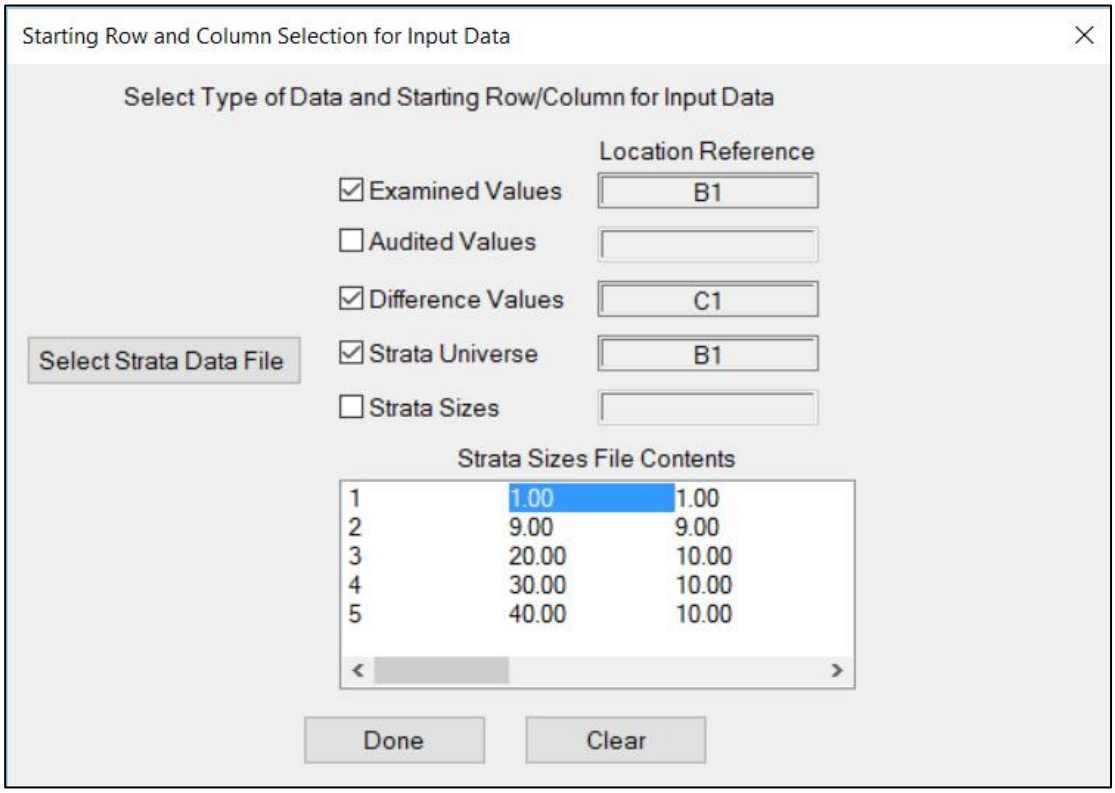

Figure 12: Text File Data Selection for Stratified Variable Appraisal

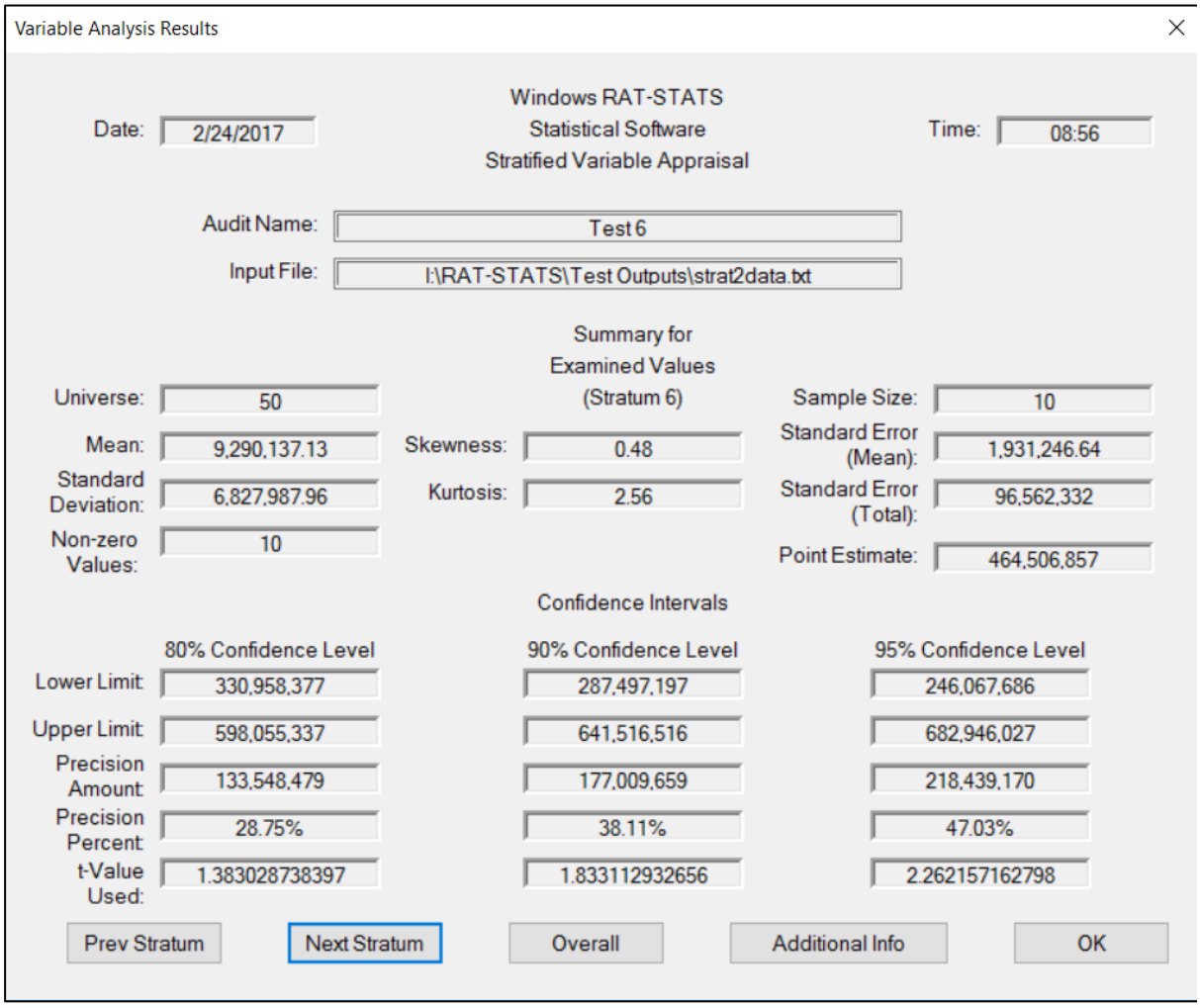

Figure 13: Stratified Variable Appraisal Results Screen

3 Differences Between RAT-STATS and RAT-STATS 2017

During testing of this solution with respect to the current implementation of RAT-STATS we came across the following differences:

3.1 GUI Differences

Although the following list of differences can be considered primarily cosmetic in nature, they do affect workflow of the end user.

3.1.1 Overall

Many of the inputs windows of the current version of RAT-STATS have a separate "EXIT" button, that terminates the program. This has been removed in favor of the standard "X" icon in the upper right hand corner of the application window. Pop-up dialog boxes have an "OK" button to acknowledge and close the dialog window.

Printing a result to a file has been relegated to clicking on the "SAVE AS.." button.

Viewing results of the screen is now selected by clicking on the "VIEW" button.

"Save As.." – When a user selects to save a file with the file extension ".xlsx" that file is automatically saved as an Excel file. Otherwise output is a plain text file.

3.2 Variable Appraisal

When examining the output of the original RAT-STATS program on screen, we discovered that the "Additional Summary Information" button results in errors displayed on the screen. Specifically, the "Standard Error (Mean)" total is not changes when you switch from the first category ("Examined") to other categories ("Audited/Difference"). Instead, that value is written to the "Standard Error (Total)" field. If the user clicks "Additional Summary Information" three times, they are back to the "Examined" pages, with now the "Standard Error (Mean)" duplicated in the "Standard Error (Total)" box, inconsistent with the value first shown. The submitted solution does not duplicate this error.

The instructions require saving output of Variable Appraisal to an Excel File, or text file that is readable by Excel. This solution generates ASCII text files in a format almost identical to the original program (there are a few spacing issues as well as differences in trailing blanks). This solution also allows for saving in an Excel file, with a separate tab for output overview, overall statistics and each stratum. The current RAT-STATs program does not generate Excel files directly, so a layout was chosen that seemed to make sense. This layout can be easily changed in the software.

Text-based input files for data and strata sizes now allow thousands separators (commas) in the numbers The files can have rows with header data. The files do not have to have the first column continuing identifiers or row numbers.

The order of columns in Excel file and text files is not specified (Examined does not have to precede Audited, and Universe size does not have to precede stratum size). In addition, the data in Excel files can be on different worksheets.

4 Licenses

The source code is copyrighted by J&J Cyber Security LLC. License to use this software is provided in compliance with the Challenge rules.

Microsoft DLLs are copyrighted by Microsoft and distributed under their End User License Agreement.

This software was developed using Microsoft Visual Studio 2015, Community Edition. This development software is freely available from Microsoft. The GUI libraries utilize the standard freely redistributable libraries (DLLs) made available with MS Windows and Visual Studio.

Input and output to MS Excel files is implemented through use of Microsoft OpenXml Software Development Software Development Kit libraries. These libraries require access using the C# language. This submission includes custom libraries, developed by J&J Cyber Security LLC, that allow C/C++ access to Excel files. These are explained in *JJCyber-Devel*. The SDK DLLs and the custom DLLs are redistributable.