

The NAEP Data Toolkit

Version 1.7

Tutorial and Reference Guide

May 23, 2001

Prepared by Educational Testing Service for the National Center of Education Statistics,
Office of Educational Research and Improvement, U.S. Department of Education

Alfred M. Rogers
Stephen Szyszkiewicz
Debra L. Kline

Educational Testing Service
Rosedale Road
Princeton, NJ 08541

**The NAEP Data Toolkit
Version 1.7
Tutorial and Reference Guide**

Table of Contents

Chapter 1: Introduction to the NAEP Data Toolkit.....	5
Chapter 2: Getting Started with the NAEP Data Toolkit.....	7
Chapter 3: Extracting NAEP Data with NAEPEX.....	11
3.1 Introduction.....	11
3.2 A Brief Tutorial.....	11
3.3 Reference Guide.....	23
Chapter 4: Using the NAEP Crosstabulation Module to Analyze NAEP Data.....	33
4.1 Opening the Crosstabulation Module.....	33
4.2 Using the Crosstabulation Module.....	35
4.3 Running the Analysis and Getting Results.....	50
Chapter 5: Using the NAEP Regression Module to Analyze NAEP Data.....	55
5.1 Opening the Regression Module.....	55
5.2 Using the Regression Module.....	57
5.3 Running the Analysis and Getting Results.....	60
5.4 NAEP Regression Tips.....	63
Appendix: Statistical Methodologies.....	65

Chapter 1: Introduction to the NAEP Data Toolkit

The NAEP Data Toolkit is a set of three programs -- NAEPEXW, the NAEP Crosstabulation Module, and the NAEP Regression Module -- designed to assist you in the management and analysis of NAEP data. The Toolkit is *not* designed as a self-contained analysis and reporting system, such as SAS or SPSS, but as a complement to those and comparable systems.

NAEPEXW provides a user-friendly interface to the NAEP secondary-use Data on Disk products by allowing you to create and manage customized extractions of data samples and to load these data into SAS or SPSS for analysis. The NAEP Crosstabulation Module and the NAEP Regression Module allow the SPSS or SAS user to construct and perform these types of analysis using statistical methodology appropriate to the design of the NAEP assessments.

NAEPEXW

Each NAEP data product contains the data for a complete set of samples and subject areas within an assessment year and assessment component (national or state). Each data file contains all of the responses, scores, weights, demographic variables, and derived variables collected for the related sample, typically numbering in the thousands of data variables. The NAEPEXW program facilitates the identification and selection of data samples and variables for study, providing you with the means to create a customized data file and the necessary information to define it to SPSS, SAS, or another user-specified data management and analysis system. For each customized data selection that you define, NAEPEXW can produce

- the extracted data file;
- a layout file containing the names, column positions, and formats of the variables in the extracted data file;
- a file of SPSS command syntax that will completely define the extract data file to SPSS;
- a file of SAS command syntax that will completely define the extract data file to SAS; and
- a file of the extracted data variables in comma-delimited format that can be loaded into a spreadsheet or database program.

NAEP Crosstabulation and Regression Modules

The NAEP Crosstabulation Module calculates a host of statistics, along with accompanying jackknife standard errors and degrees of freedom. The module accommodates the minimum N requirements used to produce NAEP results. The Crosstabulation Module enables you to perform NAEP-appropriate row-wise and column-wise t-tests, taking into consideration family size, using the *false discovery rate* (FDR) multiple comparison procedure (see the Appendix for more

information). The module provides a “mapping” feature, letting you collapse values dynamically as you structure your analyses.

The NAEP Regression Module allows you to perform regression analyses of NAEP data, with automatic calculation of correct and consistent standard errors and degrees of freedom.

These modules are designed to operate from within SPSS or as stand-alone programs if SPSS is not available. When invoked from within SPSS, each module will use the active file in the SPSS Data Editor window and put the output into the Viewer window. The modules search the lists of variable names within each data file for the appropriate variables to suggest as defaults for the weights and dependent variables in analyses.

The analyses carried out by these modules are performed using methods approved by the National Center for Education Statistics for estimating the parameters of interest and their standard errors. The results of the analyses are reproducible and consistent with published summaries of the data.

What’s Included in This Guide

Chapter 2 provides installation instructions and other information to get you started using the NAEP Data Toolkit.

Chapter 3 provides a tutorial and reference guide for the NAEPEXW data extraction program.

Chapter 4 provides a tutorial describing how to use the NAEP Crosstabulation Module to structure and carry out analyses of NAEP data; Chapter 5 provides similar information for the NAEP Regression Module.

The Appendix provides information about the methodology underlying the calculation of the crosstabulation and regression statistics and an explanation of why standard statistical procedures should not be applied to NAEP data.

Chapter 2: Getting Started with the NAEP Data Toolkit

The NAEP Data Toolkit is compatible with 386, 486, and Pentium PCs and operates under all Windows 32-bit environments (95/98/NT 4.0/2000/ME). A minimum of 4 megabytes of RAM is necessary to run the Toolkit, although, as is the case with most Windows programs, 8 megabytes of RAM is recommended. Depending on the configuration, the Toolkit will require approximately 5 megabytes of space on your hard drive. A CD-ROM drive is also required.

The following instructions assume your CD-ROM drive is letter D. Please substitute the correct letter for your CD-ROM drive if it is not D.

Installation Procedure

To install the NAEP Data Toolkit, place the NAEP Data Toolkit CD-ROM into your CD-ROM drive, and follow the prompts. If nothing happens when you place the CD-ROM in the drive, manually run the setup.exe program found in the root directory of the CD-ROM. For example, if your CD-ROM drive is designated drive D, run

```
D:\Setup.exe
```

In any case, your computer may require several restarts to complete the setup process, depending on the operating system you have on your PC and the software available on your system. This is particularly true for Windows 95/98 and NT. Once your computer restarts the installation will continue from its status prior to the reboot.

See specific operating system notes below.

Notes for Windows 95 and Windows 98 Users

Prior to installing the NAEP Data Toolkit, you must have Microsoft Internet Explorer Version 5.0 or higher installed on your PC. You can check which version of Internet Explorer (IE) you have by starting IE and then clicking on the Help menu, then clicking on About. If your version of IE is not at least 5.0 or higher, you will need to update IE to a newer version.

For your convenience we have included Internet Explorer Version 5.5 on the NAEP Data Toolkit CD. To install it run the following program from the NAEP Data Toolkit CD. If D is the letter of your CD-ROM drive, the path to install IE 5.5 SP1 is:

D:\English\Ie55sp1\i386\setup.exe

Notes for Windows NT 4.0 Users

To successfully install and execute, the NAEP Data Toolkit requires Windows NT 4.0 with Service Pack Level 5 or higher. As Windows NT is starting up, you can see which service pack you have installed. If you are unsure about which service pack you have installed, ask your systems administrator.

You can obtain the most recent service pack for Windows NT at the following web site:

<http://www.microsoft.com/ntserver/nts/downloads>

With Windows NT 4.0, the NAEP Data Toolkit must be installed by a user with administration rights for the computer.

Notes for Windows 2000 Users

With Windows 2000, the NAEP Data Toolkit must be installed by a user with administration rights for the computer.

If You Install SPSS After Installing the NAEP Data Toolkit

If SPSS is available when you install the NAEP Data Toolkit, the installation procedure described above automatically adds an item called “NAEP” to the SPSS Menu and provides the appropriate links within SPSS to the Crosstabulation and Regression Modules. If you do not have SPSS on your computer, the modules will only operate in what is referred to as “stand-alone” mode.

If you install SPSS at a later time, you can install the modules within the SPSS program by running the “SPSS Menu Extensions” program. This program is located in the NAEP Data Toolkit group in the Start menu (Start => Programs => NAEP Data Toolkit => SPSS Menu Extensions). To remove the NAEP menus from SPSS, run the “SPSS Menu Extensions” program again.

Uninstalling the NAEP Data Toolkit

To uninstall the NAEP Data Toolkit click on Start-->Settings-->Control Panel then double click Add/Remove Programs. Select NAEP Data Toolkit then click remove and follow the instructions.

You might see a dialog box announcing that a certain file could not be located. You can safely disregard this dialog box.

NOTE: If you have SPSS menu extensions installed, prior to uninstalling the NAEP Data Toolkit you should run the SPSS Menu Extensions program. (Start => Programs => NAEP Data Toolkit => SPSS Menu Extensions). This will remove the NAEP menus from SPSS. If you forget to take this step you can still remove the NAEP-specific menu system, but you will have to use the tools within SPSS to do so. Look in your SPSS Users Guide for information about Customizing Menus and Toolbars.

Chapter 3: Extracting NAEP Data with NAEPEX

3.1 Introduction

Designed for and integrated with the NAEP secondary-use data products on CD-ROM, the NAEP Data Extraction program facilitates and expedites the definition, extraction and analysis of NAEP data by secondary users. Use of the extraction program presumes a license to use the data and an understanding of the NAEP design and analysis methodology (documented in the *Data Companion* that accompanies each set of secondary-use data files).

The purpose of the NAEP secondary-use products is to provide any secondary user, whether researcher or policy maker, with *all* available data collected and derived during each assessment cycle. A typical data file can consist of hundreds, or even thousands, of variables, presenting the user with a formidable list from which to locate variables for study. The data products also include files of SPSS and SAS control statements, which will generate a complete system file for each respondent data file.

NAEPEXW works with the NAEP secondary-use data products to easily extract the information you specify. You can create, modify and manage one or more than one set of data selection criteria and generate customized data extract files, SPSS control statement files, SAS control statement files, and comma-delimited files (.CSV) for use in spreadsheet packages such as Excel or Lotus 123. All program products are saved in a designated directory on your PC workstation.

3.2 A Brief Tutorial

This section steps through a simple example to provide new users with a brief introduction to the look and feel of NAEPEXW. A more thorough discussion of the various components of NAEPEXW can be found in Section 3.3.

Registering a NAEP Data Source

Prior to beginning this tutorial, insert a NAEP secondary-use data CD into your CD-ROM drive. Every secondary-use file CD is considered a NAEP Data Source. NAEPEXW will ensure that you have a properly formatted disk inserted in your PC.

What is a NAEP Data Source? One of the key concepts associated with NAEPEXW is that of a NAEP Data Source. Each component of the NAEP assessment that has an associated secondary-use file CD is a NAEP Data Source. If the contents of a CD are copied to a hard disk

or network file server, that is also a potential NAEP Data Source. Each NAEP Data Source must be registered before it can be used. When you register a NAEP Data Source on your PC, static files are created on your hard disk that provide critical information to NAEPEXW.

To begin running the program, go to the NAEP Data Toolkit group in the Start menu and click on NAEPEXW (Start => Programs => NAEP Data Toolkit => NAEPEXW). If you have not yet registered the NAEP Data Source associated with the CD in the CD-ROM drive you will be presented with a screen that looks similar to the one shown in Figure 3-1; otherwise you will see a screen that looks similar to Figure 3-3.



Figure 3-1: NAEPEXW New CD Wizard

After you click the Next button (the only choice), NAEPEXW registers the NAEP Data Source on your PC. After you register a NAEP Data Source once you should never have to register it again. NAEPEX displays a window that shows how far along it is in registering the NAEP data Source. This window is shown in Figure 3-2.

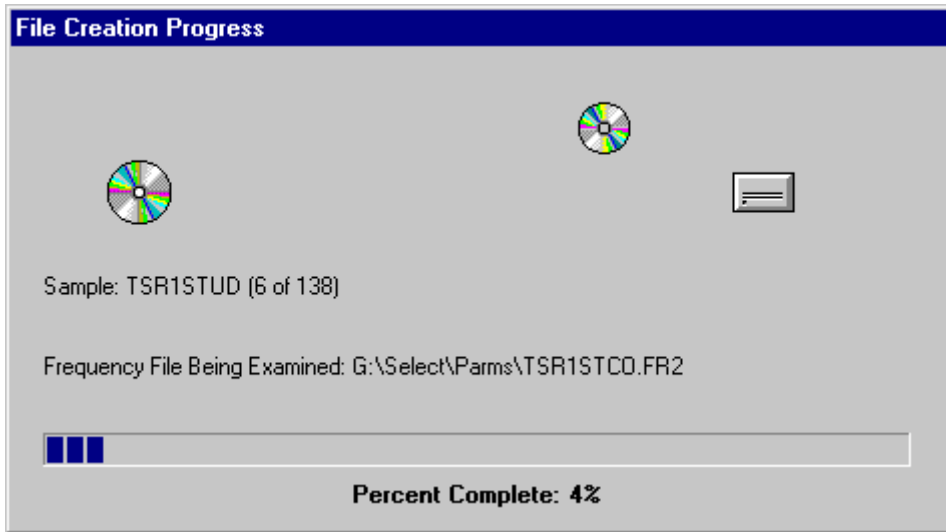


Figure 3-2: Progress of NAEP Data Source Registration

Selecting a Data Source

After registering your NAEP Data Source, you are presented with a window that looks similar to Figure 3-3. Notice that there are three tabs. The first tab is labeled "Registered Data Source." It lists all of the NAEP Data Sources that have been registered on your computer.

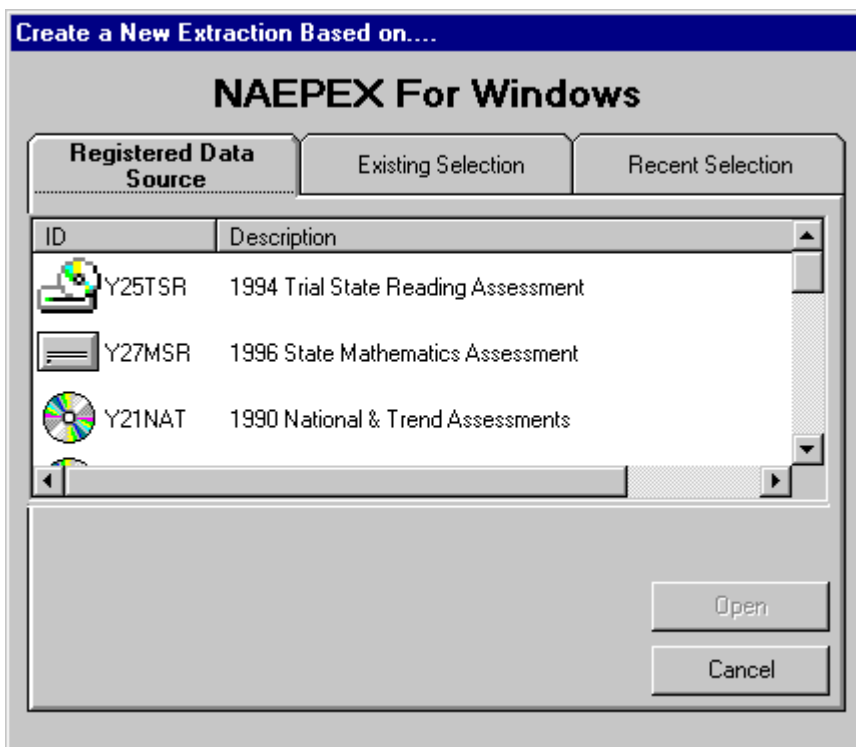





Figure 3-3: A Listing of Registered NAEP Data Sources

The NAEP Data Sources that are currently available are nearest to the top. A NAEP Data Source is considered available if either:

- the CD that defines that NAEP Data Source is in the first CD-ROM drive, or
- a secondary-use file CD has been previously copied in its entirety to a directory on an available hard drive and has been registered.

In the above example, Y25TSR and Y27MSR are both available. Y25TSR is the disk currently in the CD-ROM drive and Y27MSR is a CD-ROM disk image that was copied to an available hard disk or network drive. Y21NAT has been registered but is not currently available (because the CD-ROM drive is being used by the Y25TSR NAEP Data Source). To summarize the three symbols:

-  indicates that the NAEP data Source is available and is based on data contained on a CD-ROM
-  indicates that the NAEP Data Source is available and is based on data contained on an available hard drive or network file system
-  indicates the NAEP Data Source is not available but has been registered (data is contained on a CD-ROM)

Creating a Variable Selection Document

At the beginning of a NAEPEXW session, you can either create a new variable set or open a previously defined variable set. A “variable set” is the name associated with a group of variables, or “NAEP IDs.” Each NAEP ID identifies a particular item or parameter that was measured, assigned or calculated in the NAEP analysis. For example, *FIPS* is the NAEP ID of a variable that contains the code value for each state/jurisdiction. *SRWT01* is the NAEP ID for the first of the student replicate weights.

For the purpose of this tutorial we are going to create a new selection based on the Y25TSR data source, which contains the files for the 1994 NAEP Trial State Reading Assessment. To do so, single click Y25TSR and click the Open button. You will then be presented with a screen that looks like the one shown in Figure 3-4.

(Note: The text that appears in Figure 3-3 above and Figure 3-4 (below) for Sample may or may not match what appears on your screen, depending on which NAEP data CD you are currently using. This particular example was created using the 1994 NAEP Trial State Reading Assessment data.)

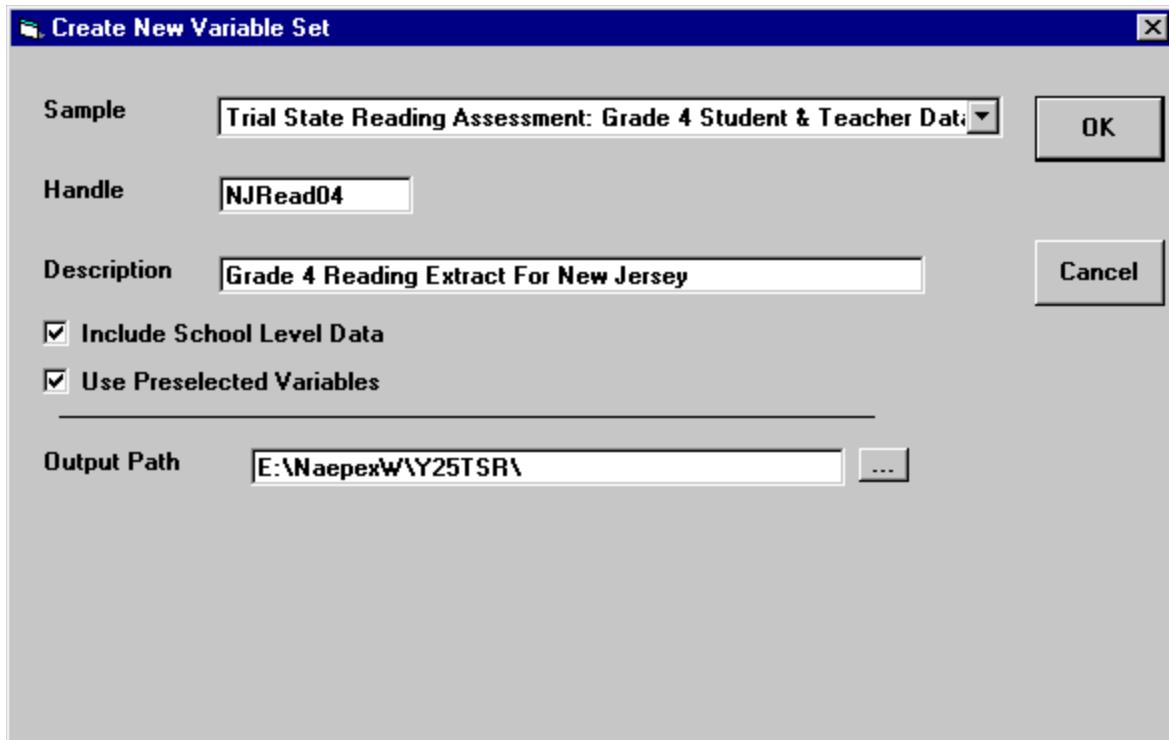


Figure 3-4: Creating a New Variable Set

A brief description of the various fields follows.

Sample

By clicking the down arrow on the Sample combo box, you will see a list of the samples that are part of the NAEP Data Source you selected previously. For now, leave the currently selected sample as the chosen sample. The text that appears in the combo box will appear on all output files.

Handle

In the Handle text box, type in:

NJRead04

A “handle” is an eight-character identifier that NAEPEXW uses to build filenames. In this particular session of NAEPEXW, all of the files that you create for this variable set will have a filename that begins with NJREAD04. For example, if later you choose to generate a SPSS syntax file it will automatically be named NJREAD04.SPS. Similarly, if you generate a SAS syntax file NAEPEXW will assign the filename of NJREAD04.SAS.

Description

In the Description text box, type in:

Grade 4 Reading Extract For New Jersey

The description is nothing more than text that NAEPEXW writes at the beginning of each file you later generate. It is there to provide you with an opportunity to write something that helps you in identifying the analysis sometime in the future.

Include School Level Data

Click the Include School Level Data check box. A checkmark will appear.

Placing a checkmark in the Include School Level Data checkbox will allow you to choose from more variables later in the process. Remember that in the NAEP design teachers and school administrators are given the opportunity to answer a series of questions about the characteristics of their particular school. By placing a checkmark in the Include School Level Data checkbox, NAEPEXW correctly matches the student to the school when it generates the various files you will use for your analysis. If you were to leave it unchecked, there would not be an opportunity to select school level variables for your analysis.

Use Preselected Variables

Leave the **Use Preselected Variables** check box in its current state (with a checkmark in it).

There are certain NAEP IDs you will use often. For example, the student replicate weights are almost always selected if you are doing a proper NAEP analysis. If you choose to use preselected variables, NAEPEXW will automatically flag a number of commonly used NAEP IDs as chosen from the hundreds that are available. If you choose not to use preselected variables, you will have to choose the entire variable set on your own later in the process.

Output Path

Either type in the name of a directory in the Output Path text box or click on the button immediately to the right of the Output Path text box to choose a directory where NAEPEXW will place all of the output files for this variable set. Figure 3-5 illustrates how you can choose a drive and directory using the mouse rather than typing the name of the directory in the textbox. (*Note: The directory name you choose is likely to be different than the one shown in Figure 3-5.*)

Next, click OK.

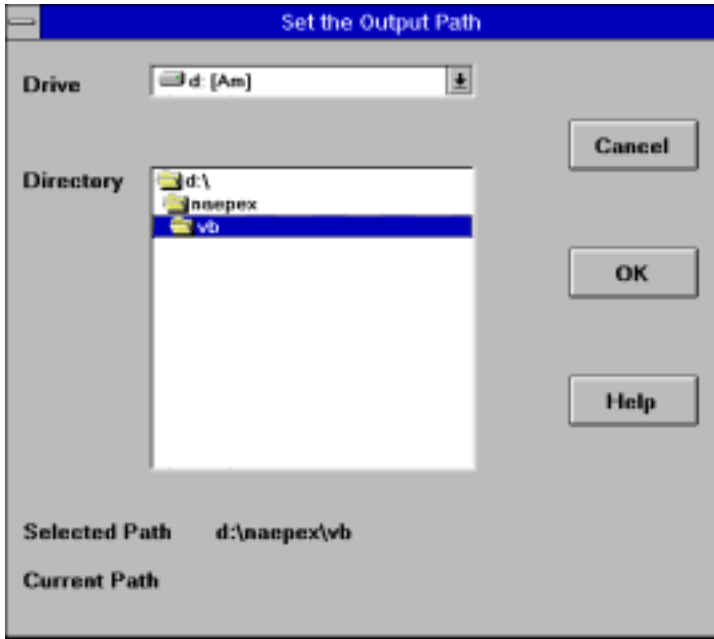


Figure 3-5: Using the Mouse to Select a Path

When you are finished with all of these steps, click the OK button on the Create New Variable Set screen. You will then be presented with a form that looks like the one in Figure 3-6. It is from this form that you will control the selection/deletion of the NAEP IDs and states that you want to include in your analysis. Since you checked the **Use Preselected Variables** check box, there are a number of commonly used NAEP IDs already included as part of your selection. The state list, however, remains empty.

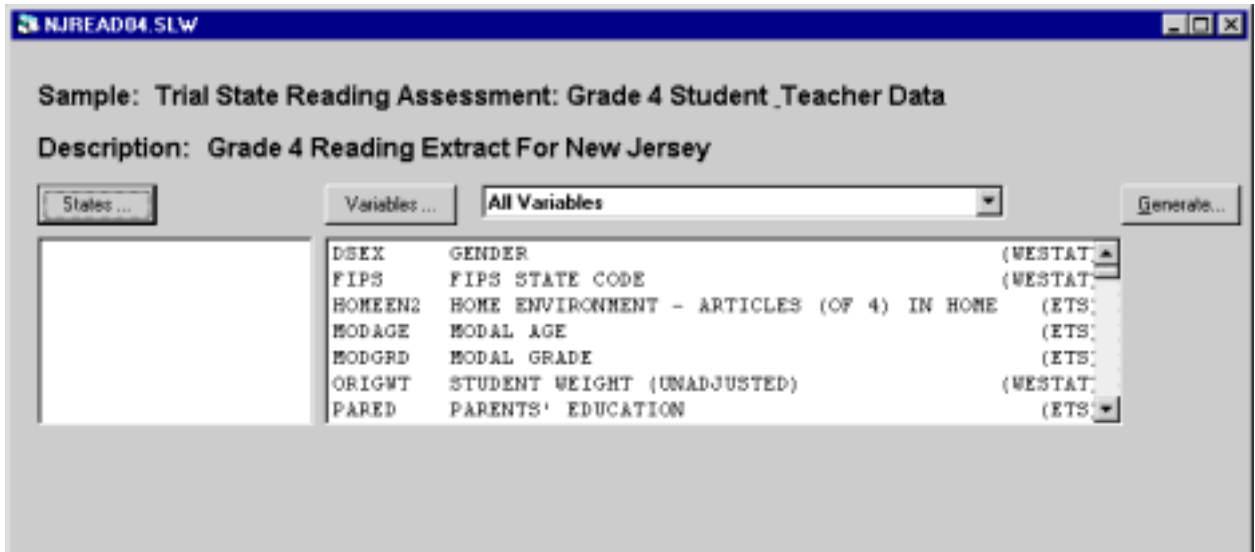


Figure 3-6: Your Current Selection

Selecting a State Sample

Click on the button labeled **States**. The screen will then display a form like the one shown in Figure 3-7.

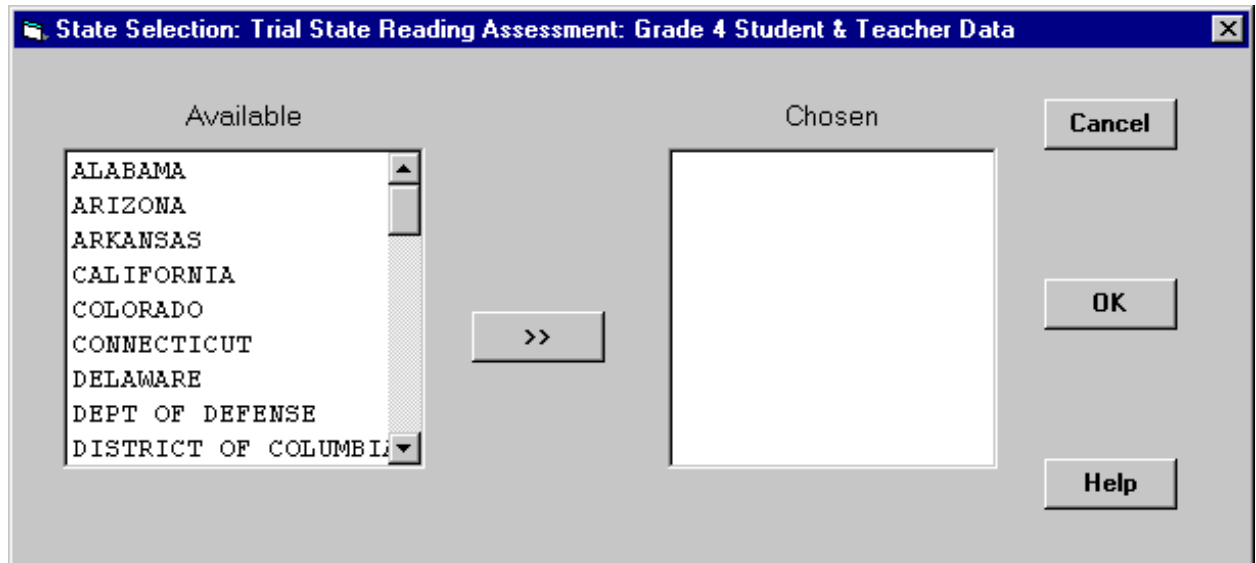


Figure 3-7: State Selection Form

To move a state from/to the **Available/Chosen** Boxes:

Click on the state you want to include and then click on the button in the middle of the form that moves the state from one box to the other; or

Double click on a state name.

To select more than one state at a time, use the Shift and Ctrl buttons in conjunction with the mouse or the arrow keys on your keyboard.

Using one of the above methods, move New Jersey to the **Chosen** list. New Jersey now appears as part of your selection. Your screen should appear like the one shown in Figure 3-8 (if you were using the 1994 NAEP Trial State Reading Student sample data). Now click the **OK** button. This will return you to the selection screen.

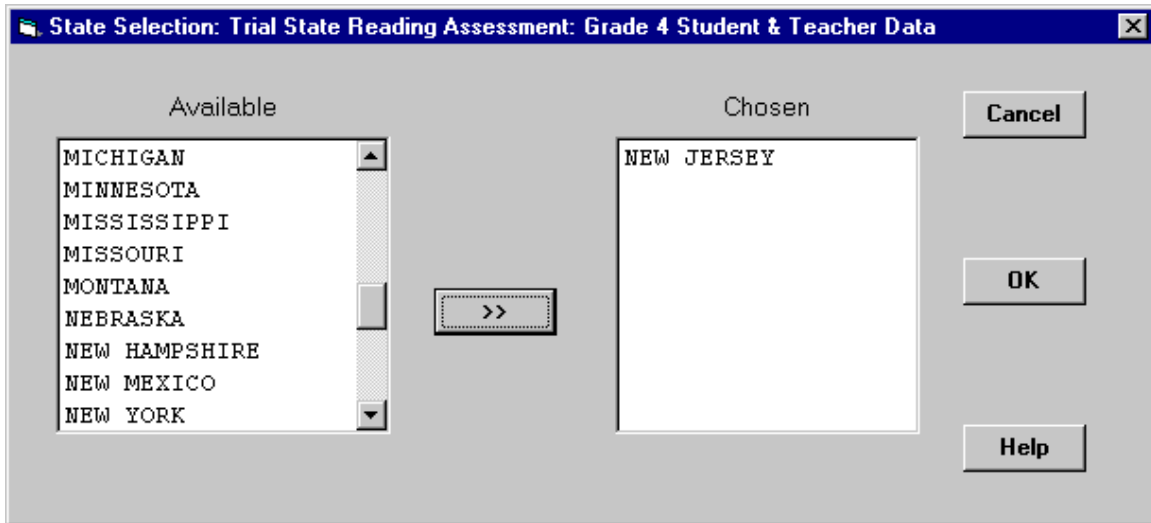


Figure 3-8: New Jersey is Now Chosen

Now click on the **Variables** button. The screen will display another dialog box, like the one shown in Figure 3-9. This form has the same functionality as the state selection form. Select the variables you want to add or remove from the selection and move them by clicking the appropriate button (Add or Remove). This form places the previously chosen variables in the **Chosen** list box and all of the remaining variables in the **Available** list box. You can determine what category each variable belongs to by clicking on the **Classification** combo box. If you change the Classification to another category and make changes to the **Chosen/Available** lists, the change is also reflected in the **All Variables** classification.

Variable Selection: Trial State Reading Assessment: Grade 4 Student & Teacher Data

Item Classification: All Variables

Available (All Variables):

ADJFAC	SCHOOL NONRESPONSE ADJUSTMENT FACTOR	(WESTAT)
ADMIN	ADMINISTRATION CODE	(BOOK COVER)
ADULTED	ADULT EDUCATION CLASSES	(QED)
AGE	ASSESSMENT AGE	
ASCHSEQ	ACROSS STATE SEQUENTIAL SCHOOL ID	(ETS)
ASSESS	ASSESSMENT CODE	(WESTAT)

Add

Chosen (All Variables):

DSEX	GENDER	(WESTAT)
FIPS	FIPS STATE CODE	(WESTAT)
HOMEEN2	HOME ENVIRONMENT - ARTICLES (OF 4) IN HOME	(ETS)
MODAGE	MODAL AGE	(ETS)
MODGRD	MODAL GRADE	(ETS)
ORIGWT	STUDENT WEIGHT (UNADJUSTED)	(WESTAT)

Buttons: OK, Cancel, Search..., Help

Figure 3-9: Variable Selection Form

Because you checked the **Use Preselected Variables** check box on the Create New Variable Set form, there are a number of commonly used NAEP IDs already included as part of your selection.

At this point you can move NAEP IDs in and out of the **Chosen** and **Available** list boxes. When you have finished choosing which variables you want to include in your analysis, click the OK button. This will return you to the selection screen.

Generating the Extract Data Files

At this point you have selected a series of NAEP IDs and states (New Jersey) for which you want to extract data. The next step is to click the **Generate** button. Do that now. First, NAEPEXW asks if you want to save your variable list to disk (as shown in Figure 3-10). Respond **Yes**.

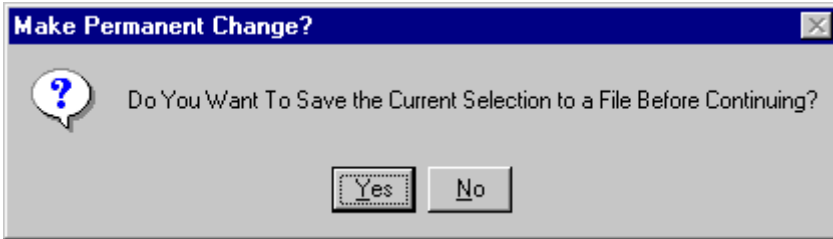


Figure 3-10: Save Your Selection to Disk For Later Retrieval

The screen will display another form like the one shown in Figure 3-11, from which you can choose the files you want to create and specify exactly where you would like them to be stored.

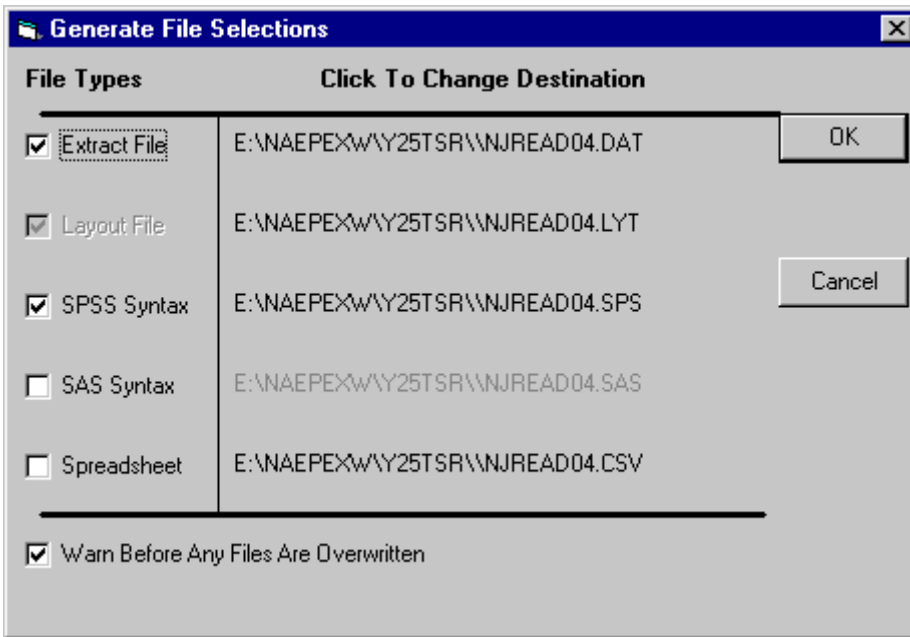


Figure 3-11: File Generation Options

For this tutorial, click the **OK** button and NAEPEXW will begin creating an Extract File, a Layout File, and an SPSS Syntax File. NAEPEXW will attempt to display all syntax (SAS and SPSS) and layout files on the screen. Extract files are not displayed because they are usually quite large and contain information that is not directly useful to the user. The SPSS Syntax and Layout files are shown in Figure 3-12.

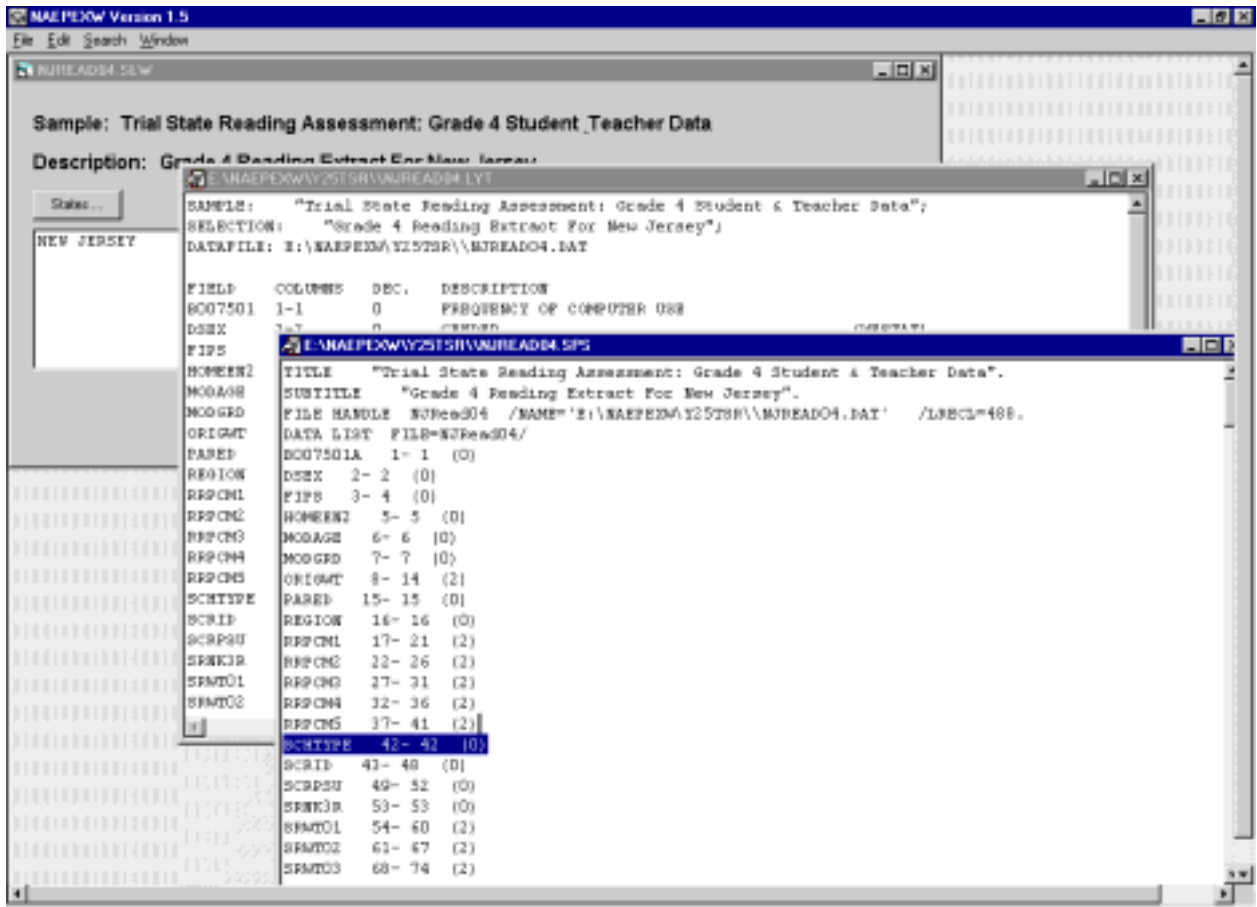


Figure 3-12: Generation Results

You can edit any of the text files on the screen as you would edit them using your favorite text editor. The clipboard is also available for easy transfer of text from the NAEPEXW program to either SAS or SPSS for Windows. When all of the files have been generated, control of the program is returned to the user. You can then examine any of the syntax files you generated (if they are not too large) immediately. You can also copy the data in any of the syntax files to the clipboard by clicking on the windows that contains the syntax file and pressing Ctrl-A, Ctrl-C. You can also use the menu system by selecting:

- Edit, Select All**
- Edit, Copy**

Once the data is on the clipboard it can be easily pasted to another Windows application such as SAS or SPSS. See your SAS/SPSS documentation for specific details.

This completes the tutorial.

3.3 Reference Guide

This section provides a screen-by-screen summary of each feature available in NAEPEXW and describes the two types of documents supported in NAEPEXW -- the Variable Selection document and the Text Editor document.

3.3.1 The Opening Screen/Variable Selection Document Menu

NAEPEXW has a Multiple Document Interface (MDI). Depending on the type of document you are currently editing, the menu system will look slightly different. There are two different types of documents in NAEPEXW. The first type of document is the Variable Selection document (VSd). Figure 3-6 shows what the menu looks like when a Variable Selection document “has the focus” or “is active.” The following is an explanation of each menu item.

VSd File Menu

To access the submenu simply click on the **File** menu or press the <ALT> key and the F key at the same time.

File New: Creates a New Selection. When you choose File New the dialog box shown in Figure 3-3 appears.

File Open: Opens an existing selection file from disk. A selection file has the default extension .SLW. For example NJREAD04.SLW is the name of the selection file created during the tutorial in Section 3.2.

File Close:	Closes the active VSd.
File Close All:	Closes all open VSd's.
File Save:	Saves the active VSd.
File Save As:	Prompts the user to save the active VSd to a different filename.
File Save All:	Saves all open VSd's.
File Exit:	Exits NAEPEXW.

The File menu also contains the names of the five most recently saved files. You can open that file by simply selecting it from the list.

VSd Edit Menu

To access the **Edit** submenu simply click on the Edit menu or press the <ALT> key and the E key at the same time.

Edit States: This launches the dialog box shown in Figure 3-7. This menu option will be disabled if a national sample was selected for analysis.

Edit Variables: This launches the dialog box shown in Figure 3-9.

VSD Generate Menu

The **Generate** menu item does not have a submenu. Clicking on **Generate** or pressing <Alt> G will launch the dialog box shown in Figure 3-11. Using that dialog box will allow you to control what types of files your VSD will create.

VSD Tools Menu

The Tools menu allows you to manipulate and register NAEP Data Sources. It has three submenus:

Register Data Source: This displays on the screen a form like the one shown in Figure 3-13 that allows you to register a NAEP data Source. If you have copied the contents of a NAEP secondary-use data CD-ROM to an available hard drive or have inserted an unregistered CD and want to register it, use this menu choice. The directory you choose should be the "virtual root" directory of the CD-ROM. For example, suppose you have copied the entire CD-ROM (including all of the subdirectories) into a directory called MYSTUFF on the C:\ drive. If you are going to register that data source you would specify C:\MyStuff as the base directory. You are also required to name the NAEP Data Source. Any string up to 8 characters long is acceptable.

You can also use the Register Data Source menu option to register or re-register a NAEP secondary use data CD-ROM, in which case you are not required to name the NAEP Data Source. The volume name of the CD is used by default.

Note: If you select "Register a Data Source Using a NAEP Secondary Use Product CD" and none is in the first available CD-ROM drive you will get an error message to that effect. If you have just placed a valid NAEP secondary-use product CD in the drive and you get that error message, you might want to select the option again. Sometimes it takes a few seconds for the computer to recognize that the CD-ROM drive has a new CD in it.

Delete Data Source: This option allows you to remove a NAEP Data Source from your system. This might be necessary to make some space on your hard drive for other applications. A NAEP Data Source copies somewhere between 10 Mb and 30 Mb of data to your system, depending on the assessment. The dialog box is shown in Figure 3-14.

Note: Removing a NAEP Data Source does not delete any extractions you may have created.

Move Data Source: This option allows you to move a NAEP Data Source from one directory to another. This might be necessary to make some space on your hard drive for other applications, but you don't want to remove the NAEP Data Source from your system. The dialog box is shown in Figure 3-15.

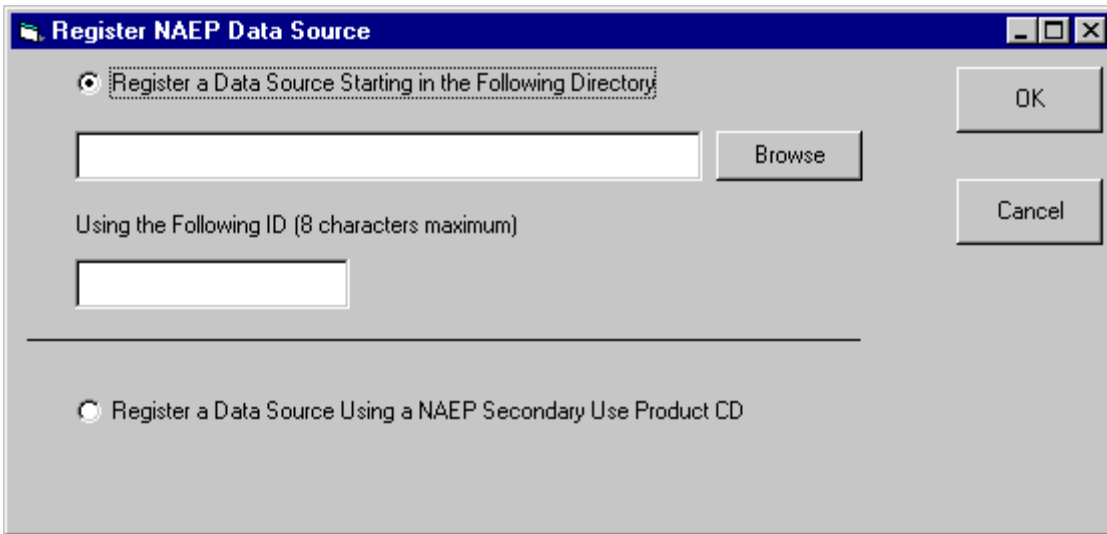


Figure 3-13: Registering a New NAEP Data Source Manually

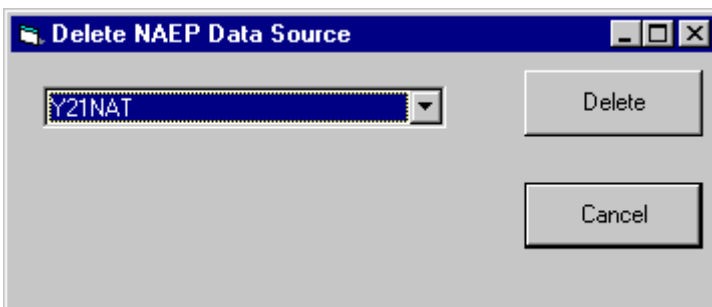


Figure 3-14: Deleting a NAEP Data Source

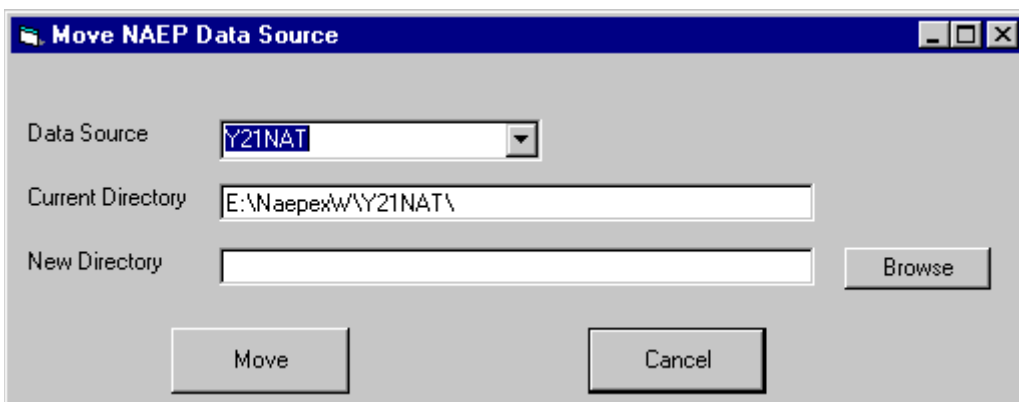


Figure 3-15: Moving a NAEP Data Source

VSD Acquire Menu

The **Acquire** menu does not have a submenu. Clicking on Acquire or pressing <Alt> A will tell NAEPEXW to search for a NAEP secondary-use data CD in your CD-ROM drive. If you have more than one CD-ROM drive, it will look in only the first one.

VSD Window Menu

To access the Window submenu simply click on the Window menu or press the <ALT> key and the W key at the same time.

- | | |
|-----------------|--|
| Window Tile: | If more than one window is visible on the screen, selecting this option will cause the screen to be divided into equal spaces and each window will appear in that space. |
| Window Cascade: | If more than one window is visible on the screen, selecting this option will cause each window to be overlapped on top of another. |

If more than one window is visible a list of all visible windows appears below the Cascade option. Selecting a window name will immediately make that window active.

VSD Help Menu

To access the Help submenu simply click the Help menu or press the <ALT> key and the H key at the same time.

- | | |
|----------------|--|
| Help Contents: | Enters the Help file at the table of contents |
| Help Search: | Examines the Help file for specific words or phrases |
| Help About: | Displays the current version of the NAEPEXW program |

Help Contents and Help Search are not currently available.

3.3.2 Text Editor Menu Choices

The other type of document that is supported by NAEPEXW is simple text. When NAEPEXW creates a SAS/SPSS syntax file or a layout file, it attempts to display the contents of those files in their own separate window. This window acts like the common Notepad program that comes with Windows. The menu is shown in Figure 3-11 (along the top of the screen).

The text editor File and Window menu choices function identically to their VSD counterparts. The File menu options apply only to the text files. The choices available from the text editor Edit menu are different from those offered by the VSD Edit menu. The Search menu is unique to the text editor menu system.

Text Editor Edit Menu

To access the Edit submenu simply click on the Window menu or press the <ALT> key and the E key at the same time.

Edit Cut:	Removes highlighted text from the text document and places it on the Windows clipboard.
Edit Copy:	Places highlighted text on the Windows clipboard without removing it from the text document.
Edit Paste:	If there are data on the clipboard this command will place the data into the text document beginning at the location of the cursor.
Edit Delete:	Deletes highlighted text from the text document but does not copy it to the Windows clipboard.
Edit Select All:	Highlights all of the text in the document.

For more information about the clipboard, see your Windows documentation.

Text Search Menu

To access the Search submenu simply click on the Window menu or press the <ALT> key and the S key at the same time.

Search Find:	Brings up a dialog box (see Figure 3-16) that allows you to type in a Search string to be found in the text document.
Search Find Next:	Allows you to look for the same string again without having to retype it.



Figure 3-16: The Search Find Dialog Box

3.3.3 The File New Dialog Box

By choosing **File New** from the VSd menu, the user is presented with a dialog box that initially appears similar to the one shown in Figure 3-3.

3.3.4 VSd Screen

The VSd screen is shown in Figure 3-6. The menu options are explained in 3.3.1.

States button: The **States** button is available only if the sample contains state-level data. If a national sample was chosen, clicking this button will have no effect. Clicking the **States** button will launch the dialog box shown in Figure 3-6.

Variables button: The **Variables** button launches the dialog box shown in Figure 3-9. Operation of that dialog box is described in detail in Section 3.3.5.

Generate button: The **Generate** button launches the dialog box shown in Figure 3-11. Operation of that dialog box is described in detail in Section 3.3.8.

Classification: You can see how various NAEP IDs are classified by clicking on the combo box that initially says **All Variables**. This acts as a filter. For example, selecting **Sampling Flags Factors and Weights** from the **Classification** combo box will result in only those NAEP IDs that are classified as such remaining visible. Output files are generated based on all of the NAEP IDs that have been selected, regardless of the classification of the NAEP IDs that are currently visible.

Help button: Clicking the **Help** button launches online help. Help is not currently available.

3.3.5 The Variable Selection Form

The Variable Selection Form is shown in Figure 3-9.

When the form is first displayed, all of the previously selected NAEP IDs (from the VSd document) appear in the **Chosen** portion of the form. All of the remaining NAEP IDs for this particular sample appear in the **Available** portion of the form. You can freely move NAEP IDs from **Chosen** to **Available** or **Available** to **Chosen** using the following methods:

- Double clicking a NAEP ID
- Single clicking a NAEP ID and clicking the **Add** or **Remove** button

Add and **Remove** function in the context of what is in the **Chosen** box. The **Add** button appears when you have clicked one or more variables from the **Available** box. The **Remove** button appears when you have clicked one or more variables from the **Chosen** box.

It is possible to select more than one NAEP ID at a time by using the Shift key or the Ctrl key as you single click a NAEP ID within the same box (this is the multiple selection feature built into the Windows operating system).

Classification: This is a filtering mechanism based on the classification of the NAEP ID. For example if you select **Sampling Flags Factors and Weights**, only those NAEP IDs classified as such are visible. Those NAEP IDs classified as **Sampling Flags Factors and Weights** that have been chosen will appear in the **Chosen** box; otherwise they will appear in the **Available** box. By selecting **All Variables** the user will see all of the NAEP IDs in alphabetical order in either the **Chosen** or **Available** lists.

OK button: Clicking the **OK** button will send the NAEP IDs in the **Chosen** dialog box back to the VSd. All of the previously selected NAEP IDs in the VSd will be replaced by those that appear in the **Chosen** list, regardless of classification.

Cancel button: Clicking the **Cancel** button returns you to the VSd that launched the dialog box without making any changes to that VSd.

Search button: Clicking the **Search** button launches the Search form. Operation of the Search form is described in Section 3.3.6. This should not be confused with the text Search menu option described in Section 3.3.2.

Help Button: Clicking the Help button launches online help. Help is not currently available.

3.3.6 The Search Form

The Search Form is shown in Figure 3-17. The intended use of this form is only to add NAEP IDs to the Variable Selection Form discussed in Section 3.3.5.

Type a word or phrase in the box labeled **Search String** and click **Begin Search**. All NAEP IDs that contain that text in either the NAEP ID itself or its associated label will appear in the **Available** or **Chosen** list boxes (depending on whether the NAEP ID was previously chosen or not). The NAEP IDs that appear in the **Chosen** list are those that matched the criteria and already were in the **Chosen** list described in Section 3.3.5.

After NAEPEXW searches the sample for the specified text and finds NAEP IDs that match the criteria, the dialog box then resembles the one shown in Figure 3-17. (The word that was searched in this case was 'computer.')

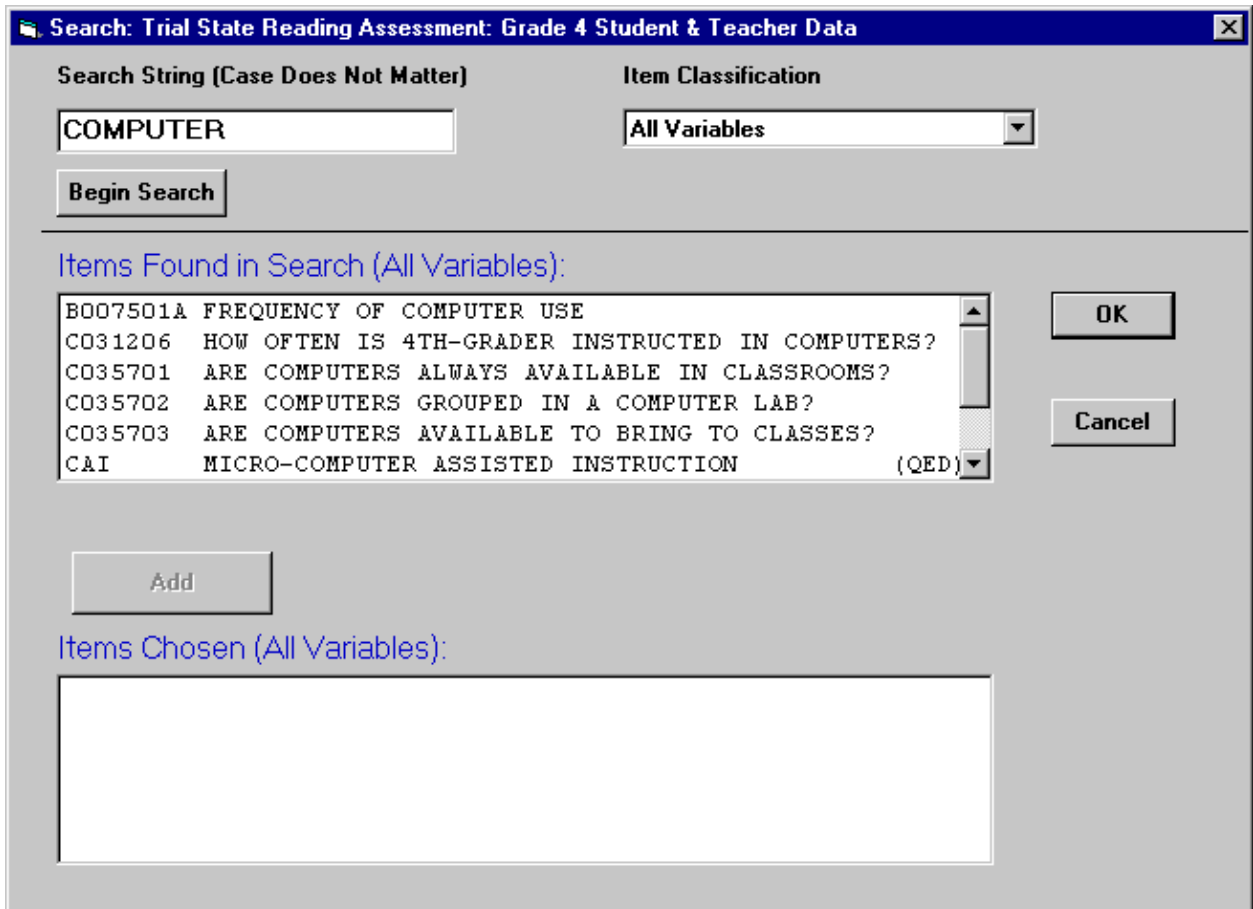


Figure 3-17: Looking for NAEP IDs Based on a Simple Word Search

OK button: Clicking the **OK** button moves the NAEP IDs in the **Chosen** list to the **Chosen** list of the Choose Variables dialog box. This does not replace what was already chosen, but adds those that were not previously chosen before the search commenced.

Cancel button: Clicking the **Cancel** button simply cancels the search and returns you to the Choose Variables dialog box.

Add button: Clicking the **Add** button moves NAEP IDs to the **Chosen** list from the **Available** list. It is only visible if one or more items in the **Available** list are highlighted.

Remove button: Clicking the **Remove** button moves NAEP IDs to the **Available** list from the **Chosen** list. It is only visible if one or more items in the **Chosen** list are highlighted.

Item Classification: This is a filtering mechanism based on the classification of the NAEP ID. For example, if you select **Sampling Flags Factors and Weights**, only those NAEP IDs that are classified as such are visible. Those NAEP IDs classified as **Sampling Flags Factors and Weights** that have been chosen will appear in the **Chosen** box; otherwise, they will appear in the **Available** box. By selecting **All Variables** you will see all of the NAEP IDs in alphabetical order in either the **Chosen** or **Available** lists (if they matched the search criteria).

Help button: Clicking the **Help** button launches online help. Help is not currently available.

3.3.7 Special Feature: Get Count

By taking advantage of a Windows feature called Pop-up Menus, unweighted counts can be displayed for each applicable NAEP ID. This feature can be activated just about anywhere a NAEP ID is visible (except in Text Editor documents). To take advantage of the feature select a NAEP ID using the left mouse button with a single click. Then, with that NAEP ID highlighted, right click the mouse button. A small menu will appear that says “Get Count.” An illustration of this is shown in Figure 3-18.

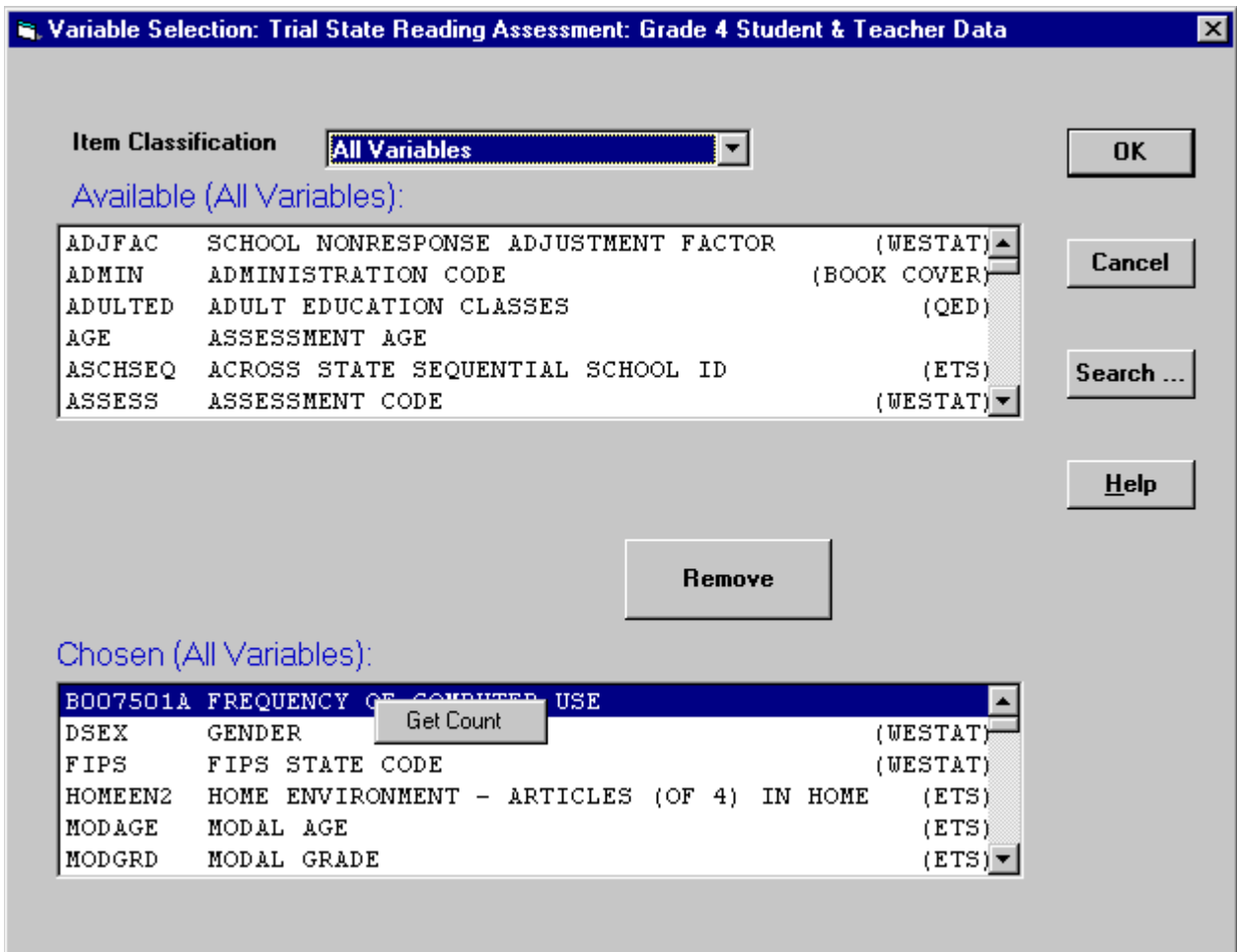


Figure 3-18: Get Count Feature

When the Get Count feature is activated, a form similar to that shown in Figure 3-19 is displayed. You can print to the default printer anything that is visible within the form or simply click OK to return to wherever Get Count was activated from in the first place. Get Count provides unweighted counts and the various categories of potential responses. For example the researcher could discover that a “yes” was coded as 1 and a “no” was coded as a 0. The count is given for each of the states that were previously selected. Get Count will not operate if there are no states selected and a national sample wasn’t chosen. Get Count will inform the user if the particular

NAEP ID, such as a replicate weight, is continuous. Count information does not exist for continuous NAEP IDs.

	NEW JERSEY
ALMOST EVERY DAY (1)	185
ONCE OR TWICE A WEEK (2)	607
ONCE OR TWICE A MTH (3)	374
NEVER OR HARDLY EVER (4)	1712
OMITTED (8)	10
MULTIPLE RESPONSE (M)	0

Figure 3-19: Get Count Results for NAEP ID B007501A

3.3.8 Generate File Selections Form

The Generate File Selections form is shown in Figure 3-11. By clicking on the various check boxes, you select which file(s) are to be generated based on the current VSd.

If you decide to generate a Data Extract File, NAEPEXW automatically creates a Layout File, thus you cannot “turn off” the Layout File generation.

If you check **Warn Before Any Files Are Overwritten**, NAEPEXW will issue a warning before any pre-existing files are overwritten.

OK button: Clicking the **OK** button will start the file generation process. If an extract file is created this may take a few minutes, depending on the number of NAEP IDs you have chosen and how many states (if applicable) are involved in the selection. NAEPEXW checks the destination drive to make sure enough space is available.

Cancel button: Clicking the **Cancel** button returns the user to the VSd without any files being generated.

Help Button: Clicking the **Help** button launches online help. Help is not currently available.

Chapter 4: Using the NAEP Crosstabulation Module to Analyze NAEP Data

4.1 Opening the NAEP Crosstabulation Module

The NAEP Crosstabulation Module can be launched in one of two ways:

- from SPSS, using the custom NAEP menu system that was established when you installed the NAEP Data Toolkit on your computer; or
- in what is called "stand-alone" mode, by clicking on the NAEP 2Way program using the familiar Windows Start menu.

Opening the NAEP Crosstabulation Module from SPSS

The custom NAEP menu from within SPSS is shown in Figure 4-1. This menu is currently operational for SPSS versions 8.0 and 9.0. When launching the NAEP Crosstabulation Module this way, the data in the currently open SPSS system file is saved to disk and then sent to the NAEP Crosstabulation Module for processing.

When launching the module in this manner, the entire data dictionary is processed, and you can begin your analysis immediately.

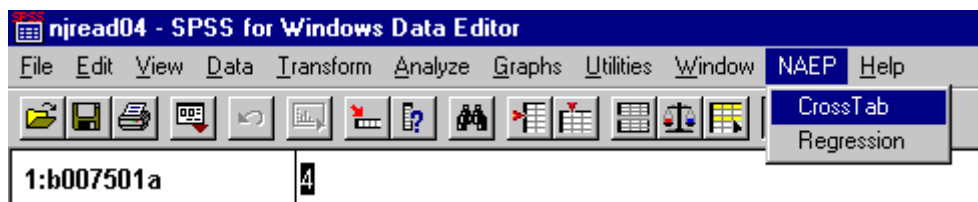


Figure 4-1: Opening the Module from SPSS

Opening the NAEP Crosstabulation Module in "Stand-Alone" Mode

It is also possible to start the NAEP Crosstabulation Module by using the familiar Windows Start button. The module can be found by clicking on Start->Programs->NAEP Data Toolkit->NAEP 2Way. This is shown in Figure 4-2. When the module is started, it asks for a .SAV (SPSS) file to open. This is shown in Figure 4-3.

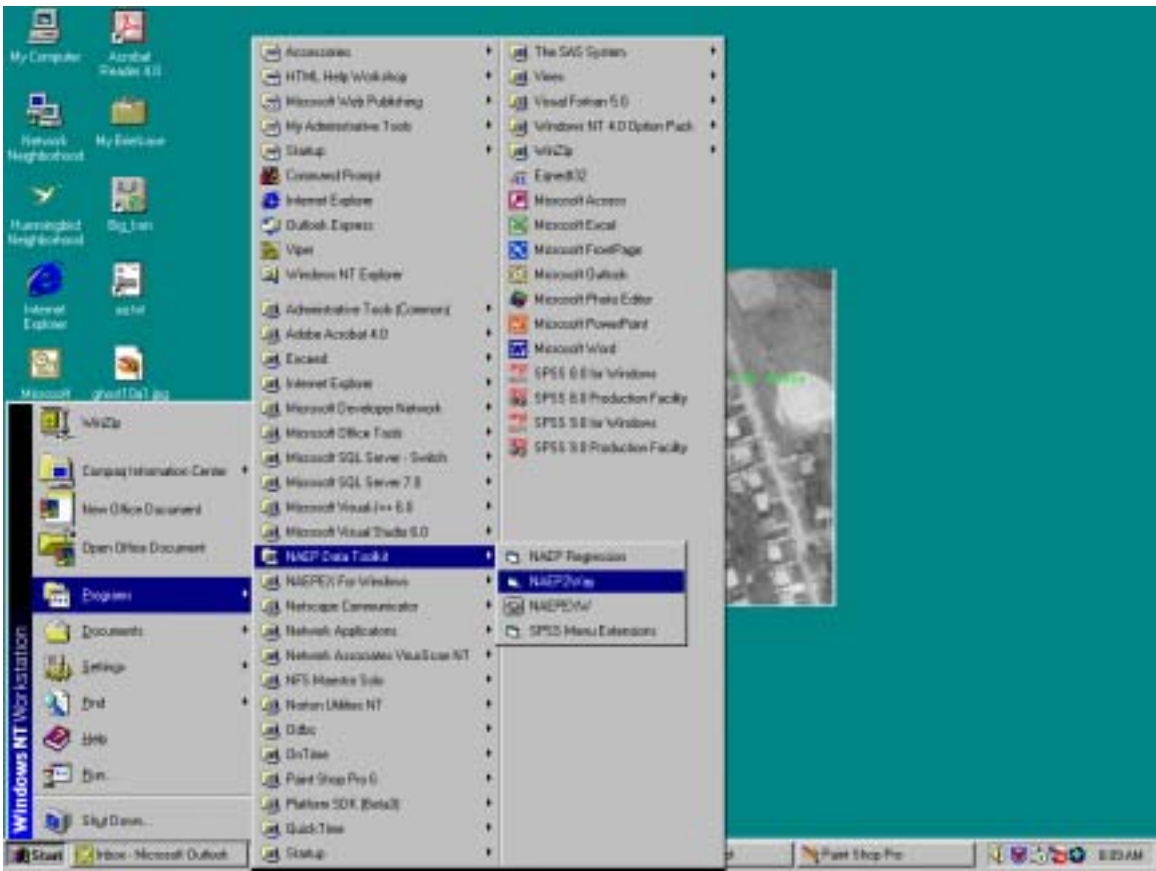


Figure 4-2: Opening the Module in Stand-Alone Mode

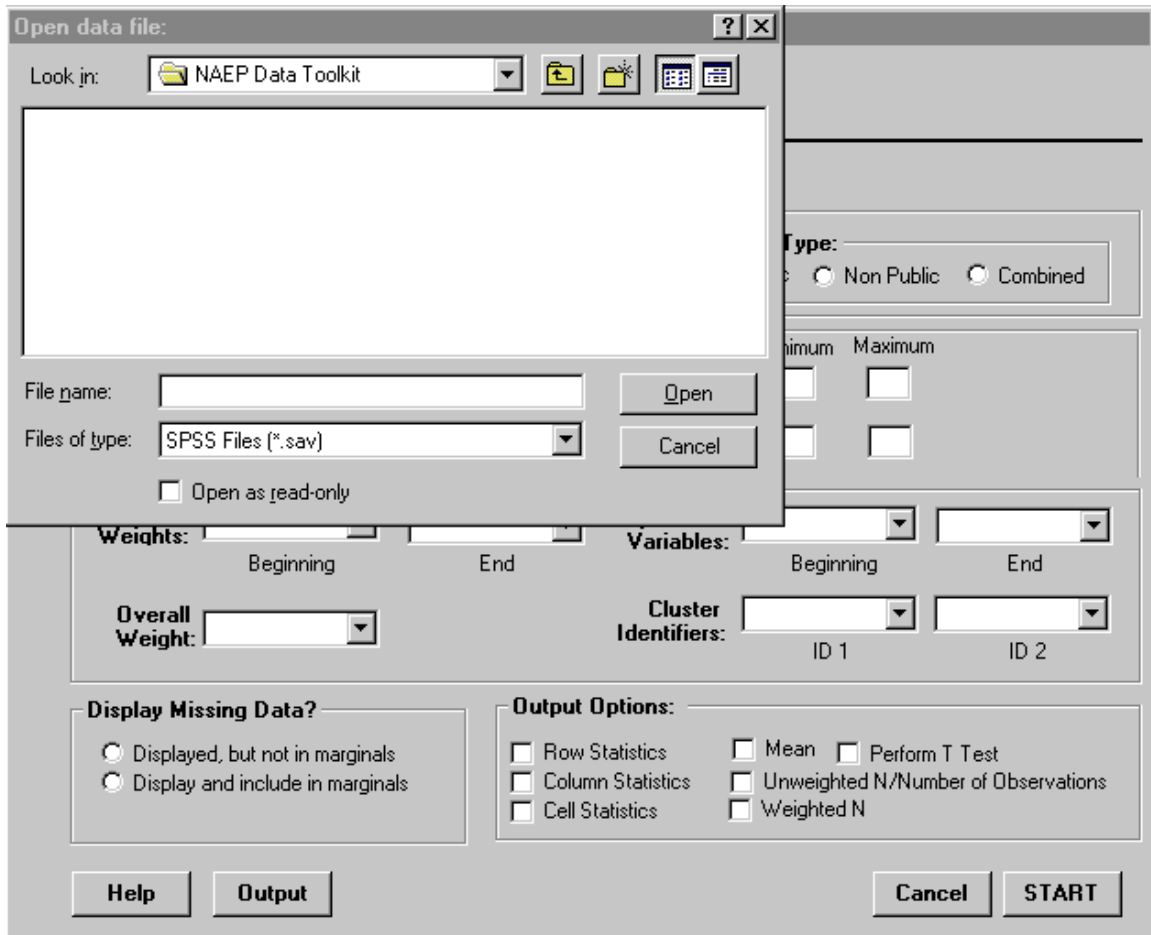


Figure 4-3: Opening a .SAV File

4.2 Using the NAEP Crosstabulation Module

Regardless of how the NAEP Crosstabulation Module was launched, it processes the data dictionary of the file and basically waits for user interaction. The user interface for the module is shown in Figure 4-4. It is divided into four main sections:

- Sample Information
- Row/Column Variable Information
- NAEP Statistical Information
- Output Options

There is also a simple menu that is available when the module is launched in Stand-Alone Mode:

Open Data File: Opens an SPSS data file

Quit Program: Quits the NAEP Crosstabulation Module

Figure 4-4: The NAEP Crosstabulation Interface

4.2.1 Sample Information

Here the user can communicate some information about the sample to the NAEP Crosstabulation Module. The first box lists the year of the sample. There are two push buttons to describe whether the sample is a state or national sample and what type of school the students in the sample attended (public/nonpublic/combined). The module uses this information to determine minimum-N requirements for reporting purposes. It is important to understand that *you must provide these three types of information* to the module – the module cannot determine or verify this information on its own. There is nothing in the data file you specified in the previous step that can uniquely identify the sample you are analyzing.

If a certain statistic does not meet the minimum-N requirements for a given sample, the results for that statistic are suppressed. There are minimum-N requirements for both the number of students and the number of schools. These requirements can vary from assessment to assessment.

Where the minimum-N requirements are enforced depends on the statistic being examined, as shown in Table 4-1.

Table 4-1: Where Minimum-N Requirements Are Enforced

Statistic	Minimum N Applies
Mean	# of students/schools in cell
Row Percentage	# of students/schools in row total
Column Percentage	# of students/schools in column total
Cell Percentage	# of students/schools in sample (total/total)

For example, the mean score for a particular cell (males whose parents' education level is less than high school) would not be reported if the total number of students in that cell was less than the minimum N for students, which is typically 62, or the total number of unique schools represented in that cell was less than the minimum N for schools, which is typically 5.

With percentages, the cell that determines the denominator must be checked to ensure that minimum-N requirements for students and schools meet the necessary requirements.

To make all of this clear, here is an example table that shows what the unweighted student counts are for a fictitious crosstabulation.

Table 4-2: A Crosstabulation to Demonstrate Minimum-N Requirements

	Col 1	Col 2	Col 3	Col 4	Total C
Row 1	13	13	13	13	52
Row 2	15	50	25	19	109
Total R	28	63	38	32	161

Assume that the minimum n for students is 62 and that the minimum N for schools has been met for all cells. Then, with the enforcement of minimum-N requirements:

- Means can be reported only for Col 2-Total R, Row 2-Total C, and Total R-Total C, because these are the only cells that have more than the minimum N for students.
- Row percentages for every cell in Row 2 can be reported because the total number of students in Row 2 is greater than 62 (it is 109)
- Row percentages for every cell in Total R can be reported because the total number of students in Total R is greater than 62 (it is 161)
- Column percentages for every cell in Col 2 can be reported because the total number of students in Col 2 is greater than 62 (it is 63)
- Column percentages for every cell in Total C can be reported because the total number of students in Total C is greater than 62 (it is 161)
- Cell percentages can be reported for every cell because the denominator for that calculation Total R-Total C is greater than 62 (it is 161).

The unweighted N statistic is never suppressed. If the unweighted N statistic is above the minimum N for students, and a particular statistic is being suppressed in the output, then the cell is being suppressed because the minimum N for schools is not being met.

The NAEP Crosstabulation Module is initially configured to enforce minimum-N requirements. There might be occasions, however, when a researcher might want to obtain a statistic for a given

cell even if the minimum-N requirements are not met. The module reads a value in the registry at the following location:

```
[HKEY_LOCAL_MACHINE\SOFTWARE\VB and VBA Program
Settings\NAEPJK>PasswordOverride]
>Password="False"
```

By changing the value of Password to "True" the module will not suppress results due to minimum-N requirements. Changing that value requires running the REGEDIT program supplied by Microsoft. Care should be taken when editing the registry, as an inadvertent change to any other setting might result in a computer that is corrupted, requiring a full installation of the underlying operating system to return it to operational status.

4.2.2 Row and Column Variable Selection

A list of all of the NAEP-IDs that are available in this data set can be seen by clicking on either the row variable or column variable box as shown in Figure 4-5 (the row variable has been clicked). It is important to note that the variables selected for row or column variables must be categorical. That is, they must have discrete values as responses. For example, DSEX (gender) is a valid row or column variable because there are two valid data values -- 1 means Male and 2 means Female. However if a plausible value is chosen (for example rrpcm1, a reading plausible value) the NAEP Crosstabulation Module would display an error message, as shown in Figure 4-6. This occurs because there is no practical way to perform a crosstabulation analysis on what the module believes to be a continuous variable. There essentially would be an infinite number of columns or rows. If a new variable is created in the data set being analyzed, that variable must contain values and value labels. If it does not, the module will assume it is a continuous variable and display the error message shown in Figure 4-6.

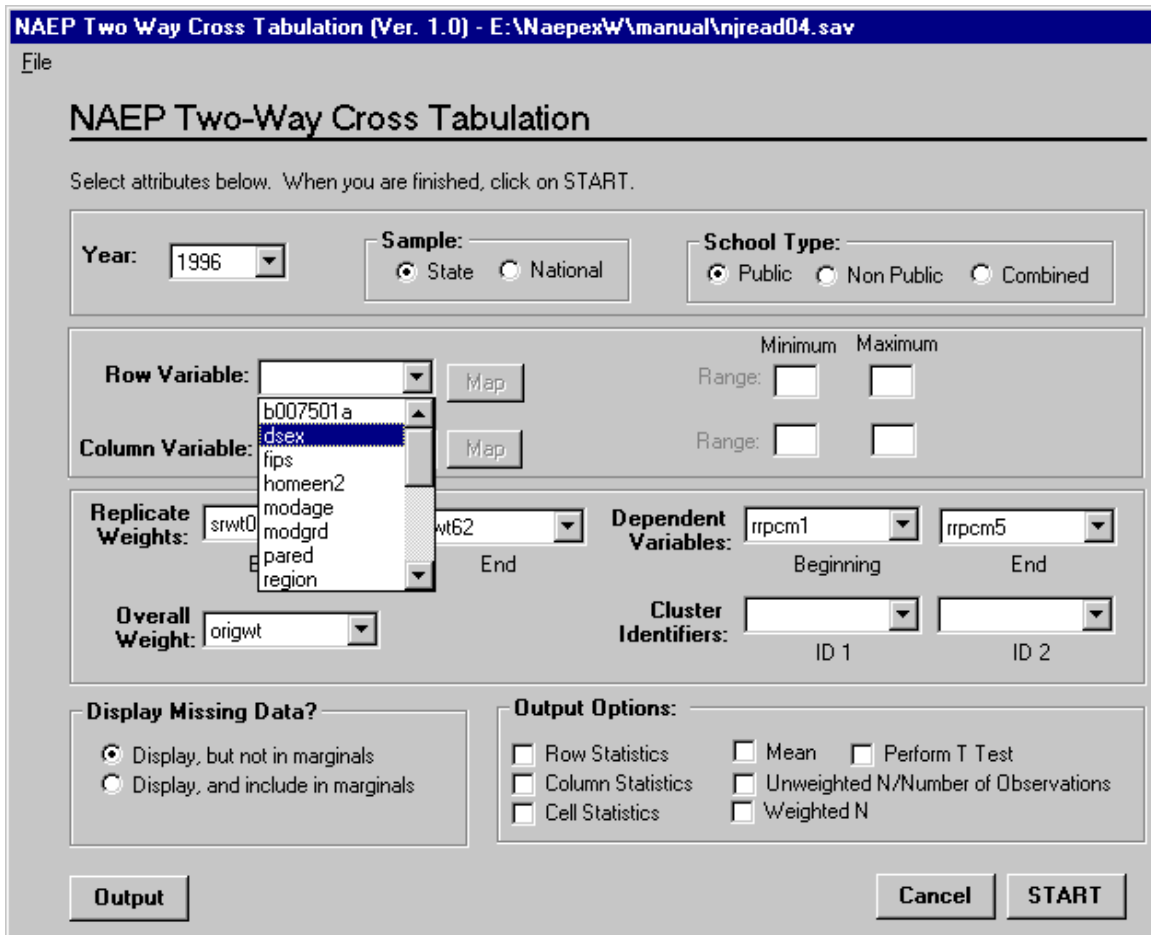


Figure 4-5: Row and Column Variable Selection



Figure 4-6: An Invalid Row or Column Variable is Detected

When a row or column variable is selected, a recommended range appears immediately in the Minimum and Maximum values. This is shown in Figure 4-7.

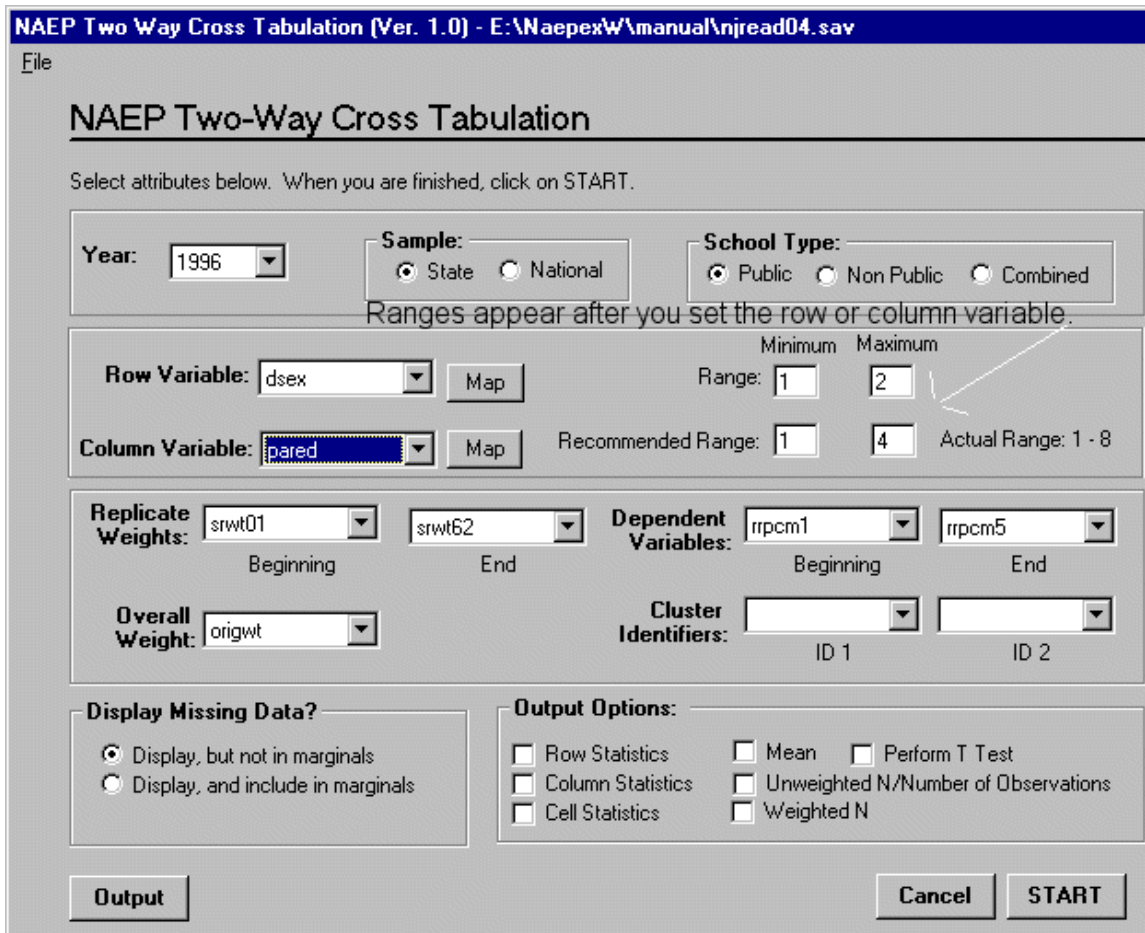


Figure 4-7: Ranges Appear for the Row and Column Variables

The variable PARED (Parental Education) has a recommended range of 1-4, but an actual range of 1-8. So what is the difference? Often in NAEP there are response categories that really are meaningless in terms of performing a relevant statistical analysis. PARED, for example, has the following values and value labels:

Value	Value Label
1	DID NOT FINISH HIGH SCHOOL
2	FINISHED HIGH SCHOOL
3	SOME COLLEGE
4	GRADUATED COLLEGE
7	I DONT KNOW
8	OMITTED

The NAEP Crosstabulation Module examines the Value Labels and makes a logical guess about what data are truly relevant and what data are not. In this case, based on the labels for values 7 and 8, it decided that the relevant range for PARED was 1-4. Specifically, the module considers the following irrelevant to an analysis: ILLEGIBLE, OFF TASK, I DON'T KNOW, NON-RATEABLE, OMITTED, NOT REACHED, and MULTIPLE RESPONSE.

The user can always change the Minimum and Maximum values to any range they please. Any response that does not fall within the recommended range is considered missing and is

accumulated into the missing cell. If, using the example above, the entire range of 1-8 were chosen, there would be two columns (or rows) that did not contain any data whatsoever (values 5 and 6).

Often the recommended range is the most appropriate range to use. There will, however, be times when the recommended range is inappropriate, or at least inconvenient. The Data Mapping feature of the NAEP Crosstabulation Module solves this problem.

Data Mapping

One of the most powerful and useful features of the NAEP Crosstabulation Module is the ability to arbitrarily map any one or more values into another user-defined value. For example, the PARED (parents' education level) variable has a valid value of 7 associated with a label of I DON'T KNOW. Normally the number of respondents who answer I DON'T KNOW to a question is very small. In the case of grade 4 students, however, this actually happens to be the most frequent response to this question, so it might make sense to combine the responses from I DON'T KNOW with the responses from another category.

Lets assume we want to combine I DON'T KNOW with FINISHED HIGH SCHOOL. That is, we would like to create the new mapping shown in Table 4-3.

Table 4-3: A Potential Mapping for PARED

Original Value	New Value
1	1
2,7	2
3	3
4	4

With this mapping, as the module is reading data from the file it is internally changing the value of a 7 to a 2 and accumulating it as such.

Another scenario that happens often is that after the first analysis is completed there exist a number of cells where the results are suppressed due to minimum-N requirements. Using the mapping features of the NAEP Crosstabulation Module it is possible to avoid this situation by combining response categories, as shown in Table 4-4.

Table 4-4: Another Potential Mapping for PARED

Original Value	New Value
1,2,7	1
3	2
4	3

In this mapping, the module will internally change values 1, 2, and 7 to a 1, value 3 to a 2, and value 4 to a 3.

To access this powerful feature for a row or column variable simply press the Map button immediately to the right of the Row or Column variable. A screen similar to the one shown in Figure 4-8 appears (for PARED). Note that the labels that appear in the box on the far left are only the labels that fall within the range given on the previous form. In Figure 4-8, there are 4

labels, because the Minimum and Maximum values were 1 and 4 respectively (Figure 4-7). If the min and max values were changed to encompass the entire range, 1 through 8, then the screen would look like the one shown in Figure 4-9.

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 4
 Foils remaining to be mapped: 4

		Collapsed Row:		Row Label:
1	DIDN'T FINISH HIGHSC	>>	<<	
2	GRAD FROM HIGH SCHOOL	>>	<<	
3	SOME ED AFTER HIGHSC	>>	<<	
4	GRAD FROM COLLEGE	>>	<<	

Automatically Map Remaining Foils

Automatically Label Rows

Help Reset Cancel OK

Figure 4-8: Recommended Range for Parents' Education Variable

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 6
Foils remaining to be mapped: 6

		Collapsed Row:		Row Label:
1	DIDN'T FINISH HIGHSC	>>	<<	
2	GRAD FROM HIGH SCHOOL	>>	<<	
3	SOME ED AFTER HIGHSC	>>	<<	
4	GRAD FROM COLLEGE	>>	<<	
7	I DON'T KNOW	>>	<<	
8	OMITTED	>>	<<	

Automatically Map Remaining Foils

Automatically Label Rows

Help Reset Cancel OK

Figure 4-9: Actual Range for Parents' Education Variable

This is a fairly complicated form, but once mastered it can provide a lot of flexibility. Here are step-by-step instructions on how to do the mapping for PARED as indicated in Table 4-4. First, click on the first value/value label pair and press the button that has the >> caption for Row 1. This will move the value/value label pair into the Row 1 list as shown in Figure 4-10. Since we want to place the original 2 and the original 7 value into this same row, click on "2 GRAD FROM HIGH SCHOOL" then press the Row 1 >> button again. It might appear that nothing has happened, but if the list for Row 1 is examined (Figure 4-11) two entries are evident, and neither entry is in the original list on the left-hand side of the screen. Repeat this step by selecting "7 I DON'T KNOW" and pressing the Row 1 >> button one last time. Row 1 is now "mapped."

It is possible to go through this process again for Row 2 and Row 3, but there is a faster way. By pressing the Automatically Map Remaining Foils button it is possible to move all of the remaining value/value label pairs to their new rows all at once. The results of this are shown in Figure 4-12.

Since the value/value label pair "8 OMITTED" was also mapped (in this case to Row 4), it is possible to remove it from the mapping scheme. To do this, simply click on the Row 4 << button. This will move the value/value label pair back to the list on the left-hand side of the screen. This is shown in Figure 4-13.

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 6
Foils remaining to be mapped: 5

		Collapsed Row:	Row Label:
2	GRAD FROM HIGHSCHOOL	Row 1 >> << 1 DIDN'T FINISH HI	
3	SOME ED AFTER HIGHSC	Row 2 >> << []	
4	GRAD FROM COLLEGE	Row 3 >> << []	
7	I DON'T KNOW	Row 4 >> << []	
8	OMITTED	Row 5 >> << []	
		Row 6 >> << []	

Automatically Map Remaining Foils

Automatically Label Rows

Help Reset Cancel OK

Figure 4-10: Data Mapping, Step 1

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 6
Foils remaining to be mapped: 4

		Collapsed Row:	Row Label:
3	SOME ED AFTER HIGHSC	Row 1 >> << 1 DIDN'T FINISH HI	
4	GRAD FROM COLLEGE	Row 2 >> << 2 DIDN'T FINISH HIGH	
7	I DON'T KNOW	Row 3 >> << 2 GRAD FROM HIGHSC	
8	OMITTED	Row 4 >> << [Empty]	
		Row 5 >> << [Empty]	
		Row 6 >> << [Empty]	

Automatically Map Remaining Foils

Automatically Label Rows

Help Reset Cancel OK

Figure 4-11: Data Mapping, Step 2

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 6
Foils remaining to be mapped: 0

		Collapsed Row:	Row Label:
Row 1	>> <<	1 DIDN'T FINISH HI	
Row 2	>> <<	3 SOME ED AFTER	
Row 3	>> <<	4 GRAD FROM COL	
Row 4	>> <<	8 OMITTED	
Row 5	>> <<		
Row 6	>> <<		

Automatically Map Remaining Foils

Automatically Label Rows

Help Reset Cancel OK

Figure 4-12: Data Mapping, Step 3

NAEP Two Way Cross Tab

Please map the foils for this variable by first selecting a foil in the "Foils" listbox on the left. Then click a ">>" button to move the foil into a row. You may map more than one foil to a row. You can also use the "<<" button to move a foil you have mapped back to the original list for remapping.

Column Variable: pared

Total # of foils: 6
Foils remaining to be mapped: 1

		Collapsed Row:	Row Label:
8 OMITTED	Row 1	>> << 1 DIDN'T FINISH HI	
	Row 2	>> << 3 SOME ED AFTER	
	Row 3	>> << 4 GRAD FROM COL	
	Row 4	>> << []	
	Row 5	>> << []	
	Row 6	>> << []	

Automatically Map Remaining Foils


Automatically Label Rows

Help Reset Cancel OK

Figure 4-13: Data Mapping, Step 4

To return back to the main form, press the OK button. The program provides a warning that not all of the value/value labels have been mapped, and the unmapped values will be considered missing. This is shown in Figure 4-14. Since that is the intended effect, simply press the Yes button and continue on. The NAEP Crosstabulation Module creates default labels for all mapped values. You can create your own labels by using the text boxes on the right hand side of the mapping screen. If you don't create labels, the module will warn you that it is about to create labels for you (Figure 4-15). Click No to allow the module to generate default labels.

Treat as missing?

 One or more foils have not been mapped. These will be treated as missing. Are you sure you wish to treat them as missing?

Yes No

Figure 4-14: Unmapped Values are Detected

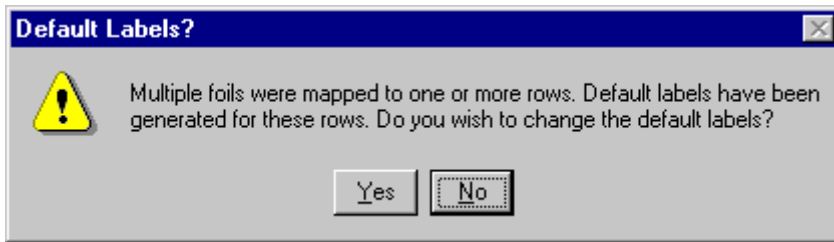


Figure 4-15: Select or Change Default Labels

When the main form pops back up, the Minimum and Maximum values for PARED will be dimmed, indicating that these data cannot be changed unless the Map button is pressed again and a new mapping is specified. It is possible to remap the features by pressing the Map button again.

NOTE: The mapping feature does not change any of the original data. The module dynamically maps the data as it reads them from the data source. The original data remain unchanged.

4.2.3 NAEP Statistical Information

Weights

There are two types of weights that are necessary to properly use the NAEP Crosstabulation Module: replicate weights and the overall weight.

Based on the data dictionary, the NAEP Crosstabulation Module will make an informed guess at what the weight values should be for a given data set and select them by default. The program does, however, let you change these selections. In general, you should use the default selections unless you have good reason not to. For example, you might have created a data file with more than one set of replicate and overall weights and it is necessary to select a specific set of weights for the sample you are analyzing.

Plausible Values

Like the weights, the module makes an attempt to figure out what NAEP IDs are the plausible values and makes them the default selection. In general it is not necessary to change these selections, but it could be possible, if the sample you are using contains more than one set of plausible values. This might happen if scores are being reported for more than one scale for a given subject, and the plausible values from more than one scale were included as part of the selection when the data set was created.

Cluster Variables

Like the weights and plausible values, the module makes an attempt to figure out which NAEP IDs are the cluster variables. Cluster variables are used by the NAEP Crosstabulation Module to count the number of unique schools/clusters in any given cell. Both cluster variables are mandatory. If there is only one cluster variable in the data set, make the second cluster variable the same as the first.

4.2.4 Output Options

The NAEP Crosstabulation Module handles missing data in one of two ways. To be clear, any data value is considered "missing" if it is not within the Minimum and Maximum value range

specified for that variable, or if it is not within the mapped range defined by the mapping screen. Essentially, the module has the flexibility to include the missing data as part of the total or not. In general, missing data should not be included as part of the total, and that is the default.

For each cell, a number of statistics can be calculated, including

- Row percentages (including standard errors and degrees of freedom)
- Column percentages (including standard errors and degrees of freedom)
- Cell percentages (including standard errors and degrees of freedom)
- Mean (including standard errors and degrees of freedom)
- Unweighted N
- Weighted N

NAEP-Appropriate Significance Testing

When a check is placed in the **Perform T Test** box, the module will automatically perform significance testing on the means in each cell. Two separate families of tests will be performed: row-wise and column-wise. If there were three rows and five columns, a "row-wise" test would include the following tests:

Row 1 Column 1 vs. Row 1 Column 2
Row 1 Column 1 vs. Row 1 Column 3
Row 1 Column 1 vs. Row 1 Column 4
Row 1 Column 1 vs. Row 1 Column 5
Row 1 Column 2 vs. Row 1 Column 3
Row 1 Column 2 vs. Row 1 Column 4
Row 1 Column 2 vs. Row 1 Column 5
Row 1 Column 3 vs. Row 1 Column 4
Row 1 Column 3 vs. Row 1 Column 5
Row 1 Column 4 vs. Row 1 Column 5
Row 2 Column 1 vs. Row 1 Column 2
Row 2 Column 1 vs. Row 2 Column 3
Row 2 Column 1 vs. Row 2 Column 4
Row 2 Column 1 vs. Row 2 Column 5
Row 2 Column 2 vs. Row 2 Column 3
Row 2 Column 2 vs. Row 2 Column 4
Row 2 Column 2 vs. Row 2 Column 5
Row 2 Column 3 vs. Row 2 Column 4
Row 2 Column 3 vs. Row 2 Column 5
Row 2 Column 4 vs. Row 2 Column 5
Row 3 Column 1 vs. Row 3 Column 2
Row 3 Column 1 vs. Row 3 Column 3
Row 3 Column 1 vs. Row 3 Column 4
Row 3 Column 1 vs. Row 3 Column 5
Row 3 Column 2 vs. Row 3 Column 3
Row 3 Column 2 vs. Row 3 Column 4
Row 3 Column 2 vs. Row 3 Column 5
Row 3 Column 3 vs. Row 3 Column 4
Row 3 Column 3 vs. Row 3 Column 5
Row 3 Column 4 vs. Row 3 Column 5

A "column-wise" test would include the following comparisons:

Row 1 Column 1 vs. Row 2 Column 1
Row 1 Column 1 vs. Row 3 Column 1
Row 2 Column 1 vs. Row 3 Column 1
Row 1 Column 2 vs. Row 2 Column 2
Row 1 Column 2 vs. Row 3 Column 2
Row 2 Column 2 vs. Row 3 Column 2
Row 1 Column 3 vs. Row 2 Column 3
Row 1 Column 3 vs. Row 3 Column 3
Row 2 Column 3 vs. Row 3 Column 3
Row 1 Column 4 vs. Row 2 Column 4
Row 1 Column 4 vs. Row 3 Column 4
Row 2 Column 4 vs. Row 3 Column 4
Row 1 Column 5 vs. Row 2 Column 5
Row 1 Column 5 vs. Row 3 Column 5
Row 2 Column 5 vs. Row 3 Column 5

T-test Specifications

The t-test is two-tailed and the significance level begins at 0.05 and is adjusted according to the FDR multiple comparison methodology. The t-test is carried out only for the means; if means are not selected, the t-tests are not performed. Totals (row or column) are not included in the family size calculation and are not included as part of any t-test. If the missing data are included as part of the total, the missing data are also included in the t-test.

NOTE: Caution is advised when choosing to perform a t-test involving a variable that has more than 7 or so possible values. As the small example above illustrates, the number of tests that get performed increases significantly as the number of rows/columns increases. This has a tremendous effect on the output file size and calculation time. The module will perform the tests; it just might take a good long while to finish them and print the results.

4.3 Running the Analysis and Getting Results

When all of the data have been properly entered, a NAEP Crosstabulation can be calculated. To start the process, simply click the Start button (**not** the Windows Start button) in the lower right-hand corner of the module screen. Once the analysis begins you are kept aware of its progress and can cancel it at anytime by pressing the Cancel button.

If the NAEP Crosstabulation Module was launched from SPSS, the process will begin immediately. If, however, the module was started in "Stand-Alone" mode, a prompt is displayed asking for a name for the output file (Figure 4-16). A default file extension of .TXT will be assigned to the file, which will be saved as a standard text file that is readable by any editor.

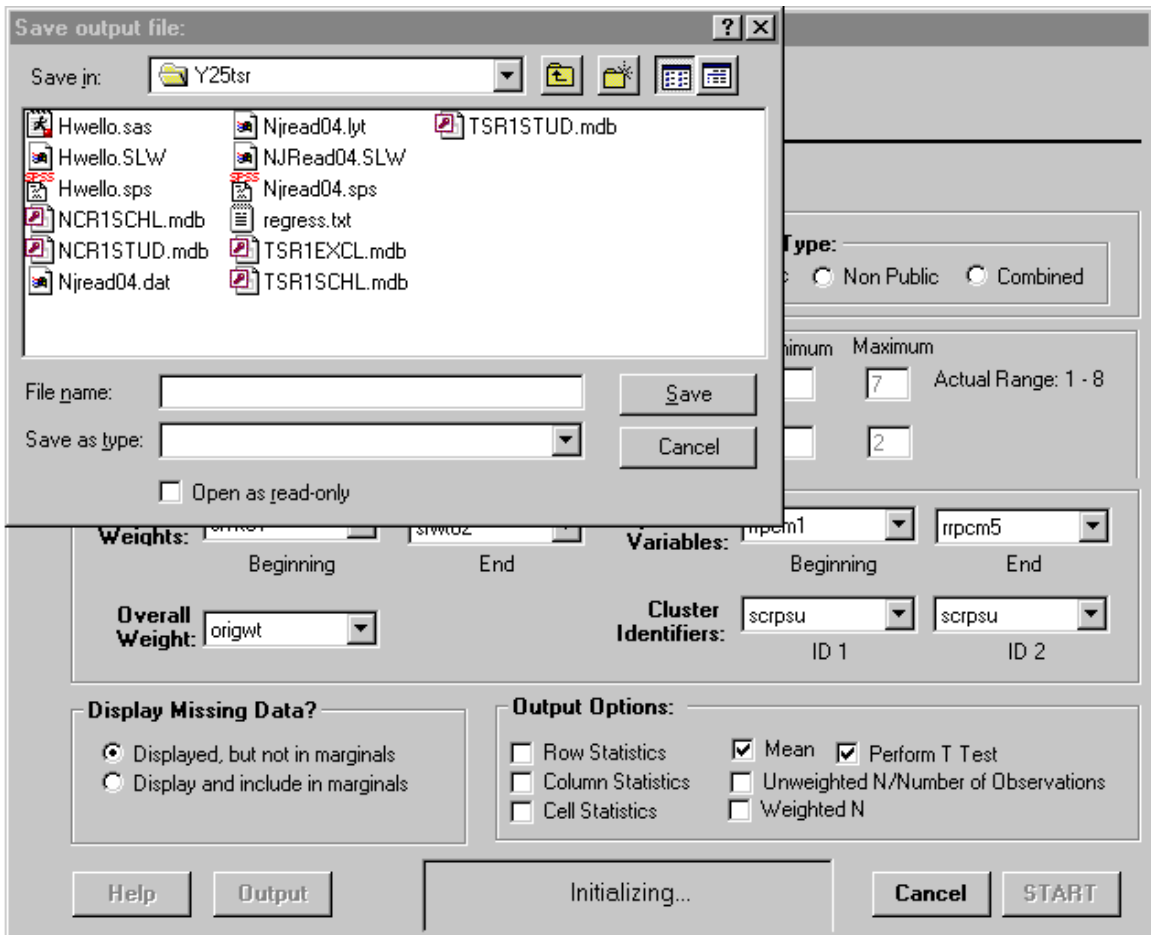


Figure 4-16: Saving the Output in Stand-Alone Mode

If the NAEP Crosstabulation Module was launched from the NAEP menu in SPSS, the output is sent immediately to an SPSS Output window (Figure 4-17). The module is then immediately unloaded.

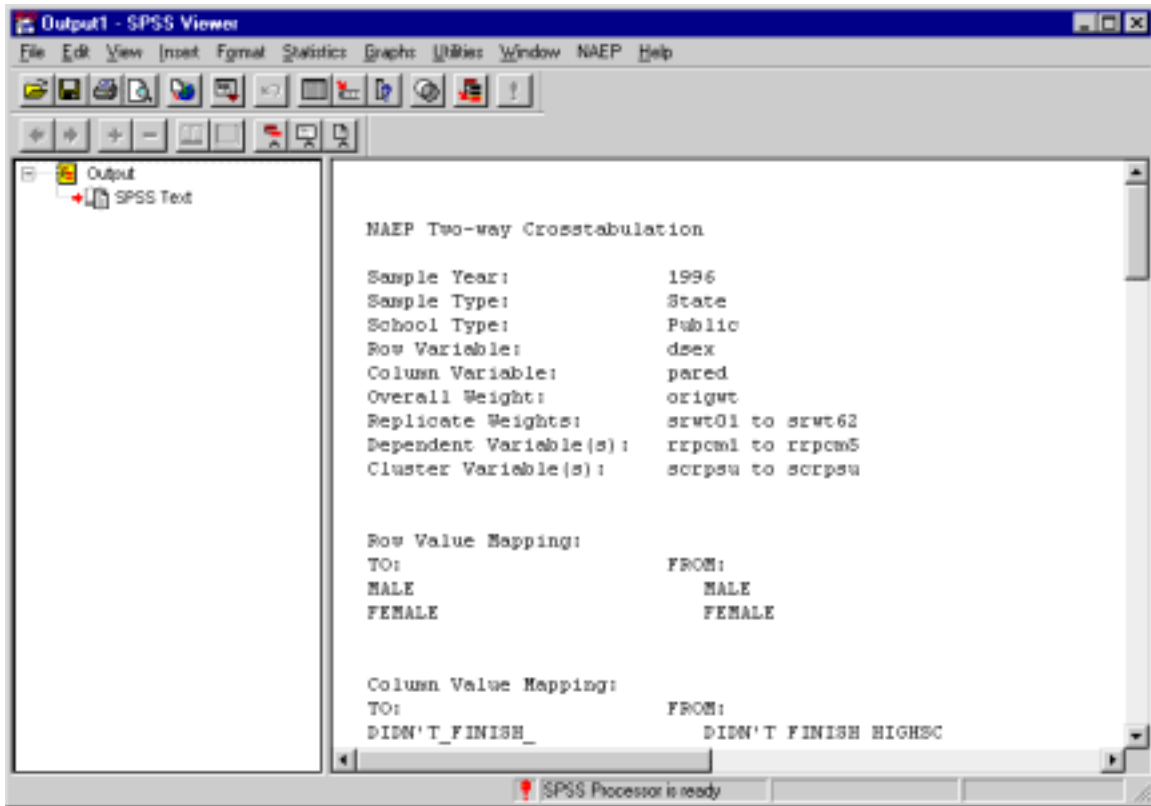


Figure 4-17: Results Displayed in SPSS Output Window

If the NAEP Crosstabulation Module was launched in “Stand-Alone” mode, the output is sent to the output file and a copy of the file is displayed in the module (as shown in Figure 4-18). Pressing OK will close the output box; then pressing cancel will close the NAEP Crosstabulation Module program.

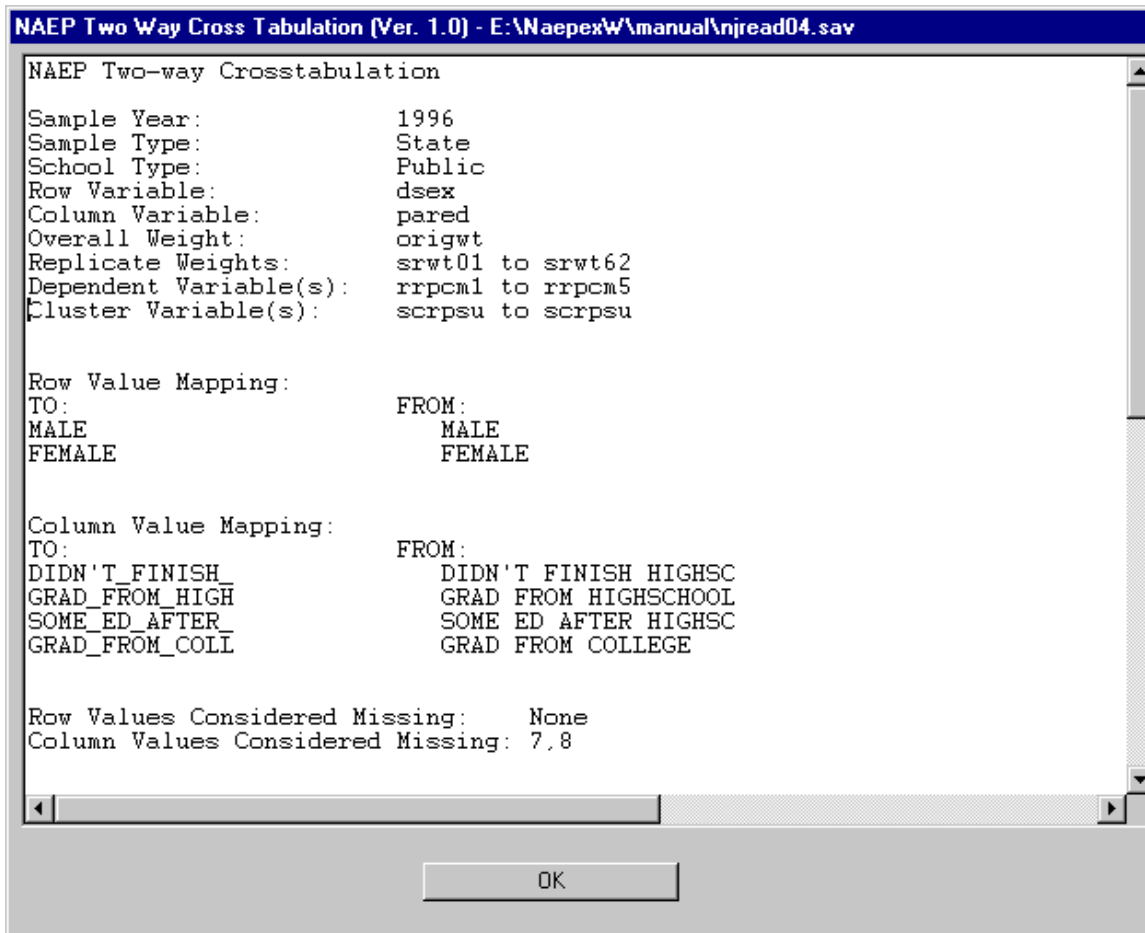


Figure 4-18: Results Displayed in Stand-Alone Module Window

Chapter 5: Using the NAEP Regression Module to Analyze NAEP Data

5.1 Opening the NAEP Regression Module

The NAEP Regression Module can be launched in one of two ways:

- from SPSS, using the custom NAEP menu system that was established when you installed the NAEP Data Toolkit on your computer; or
- in what is called “stand-alone” mode, by clicking on the NAEP Regression program using the familiar Windows Start menu.

Starting the Regression Module from SPSS

The custom NAEP menu from within SPSS is shown in Figure 5-1. This menu is currently operational for SPSS versions 8.0 and 9.0. When launching the NAEP Regression Module in this way, the data in the currently open SPSS system file is saved to disk and then sent to the NAEP Regression Module for processing.

When launching the module in this manner, the entire data dictionary is processed and you can begin your analysis immediately.

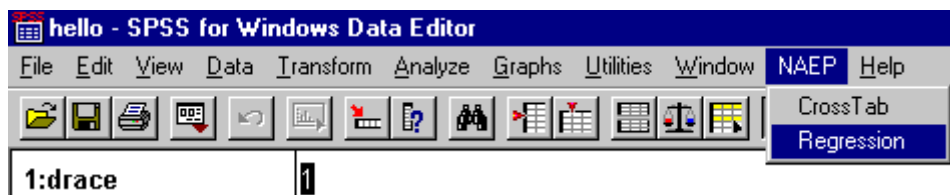


Figure 5-1: Opening the Module from SPSS

Opening the Regression Module in “Stand-Alone” Mode

It is also possible to start the NAEP Regression Module by using the familiar Windows Start button. The module can be found by clicking on Start->Programs->NAEP Data Toolkit->NAEP Regression (Figure 5-2). When the module is started, it asks for a .SAV (SPSS) file to open (Figure 5-3).

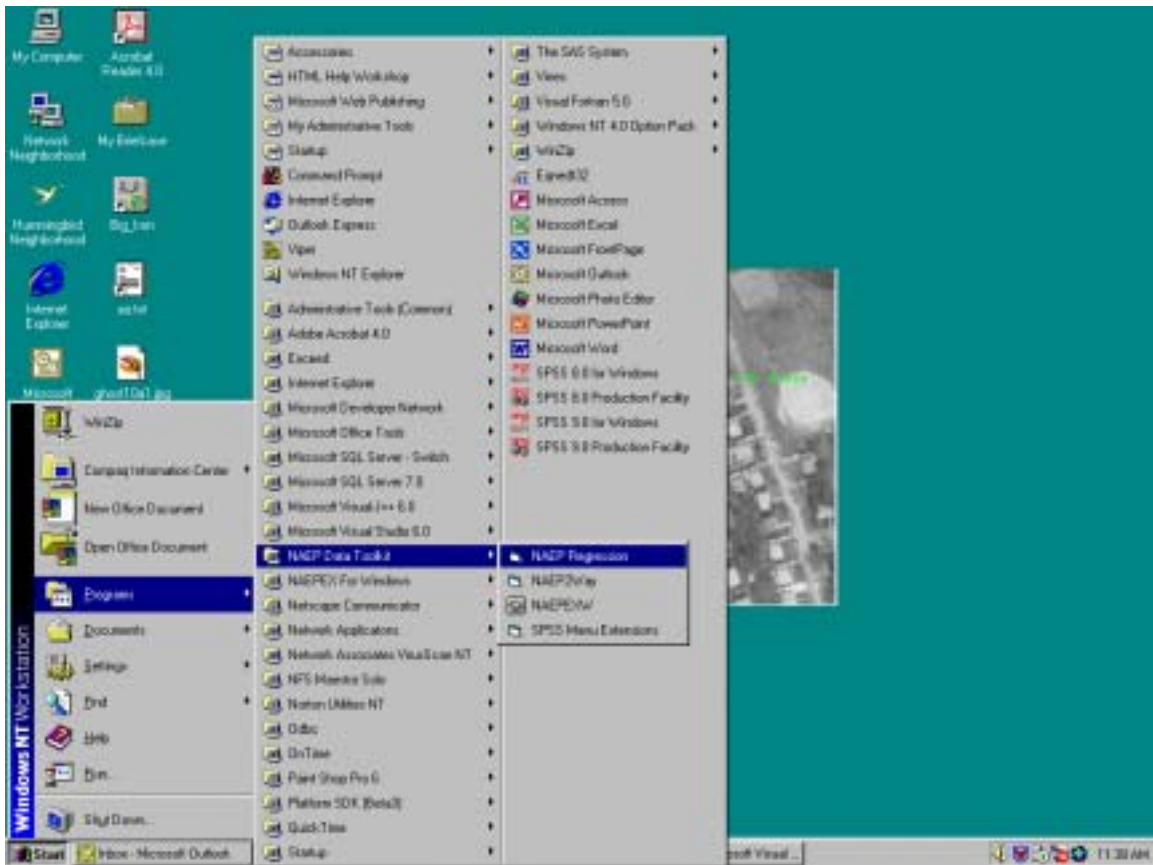


Figure 5-2: Opening the Module in Stand-Alone Mode

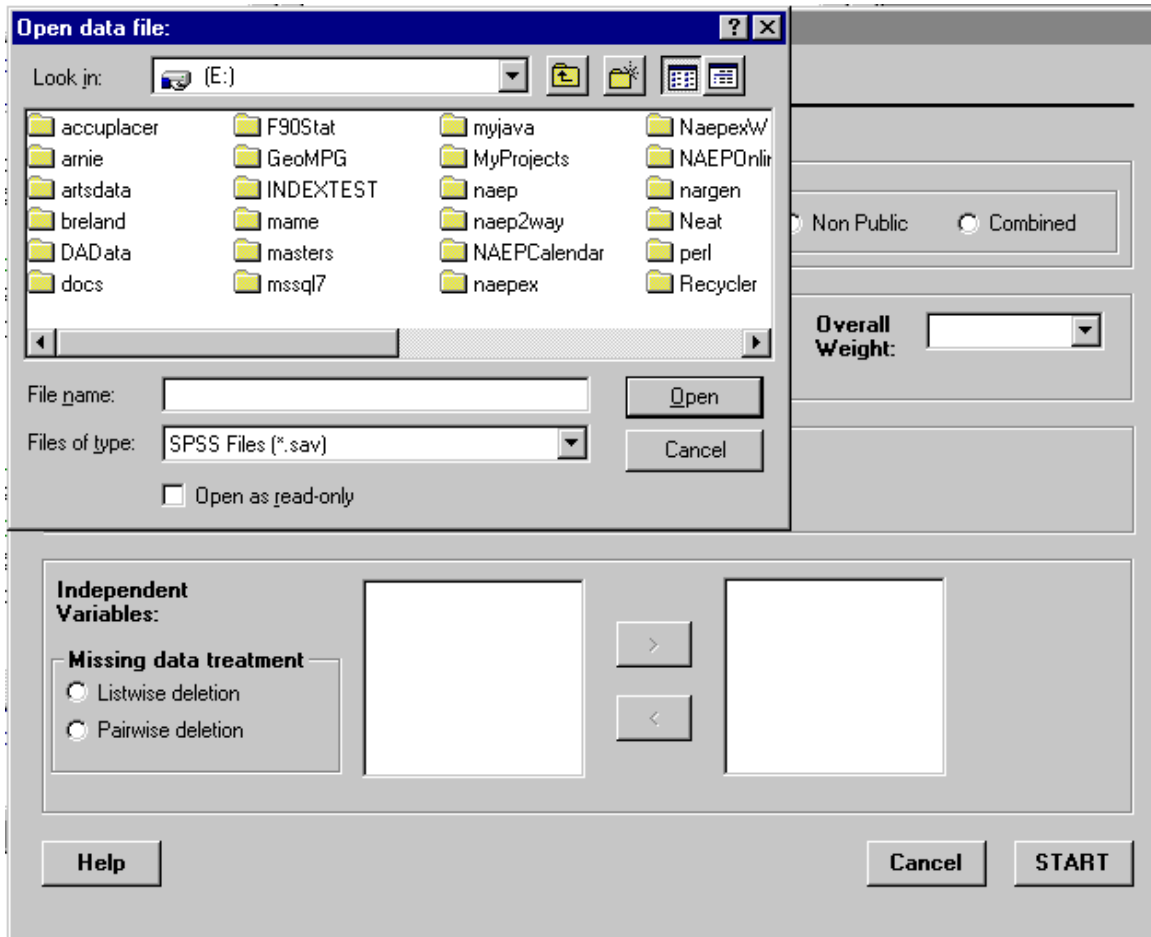


Figure 5-3: Opening a .SAV File

5.2 Using the NAEP Regression Module

Regardless of how the NAEP Regression Module was launched, it processes the data dictionary of the file and basically waits for user interaction. The user interface for the module is shown in Figure 5-4. It is divided into four main sections:

- Sample Information
- Weight Information
- Dependent Variables
- Independent Variables

There is also a simple menu that is available when the module is launched in stand-alone mode:

Open Data File: Opens an SPSS data file.

Quit Program: Quits the NAEP Regression Module

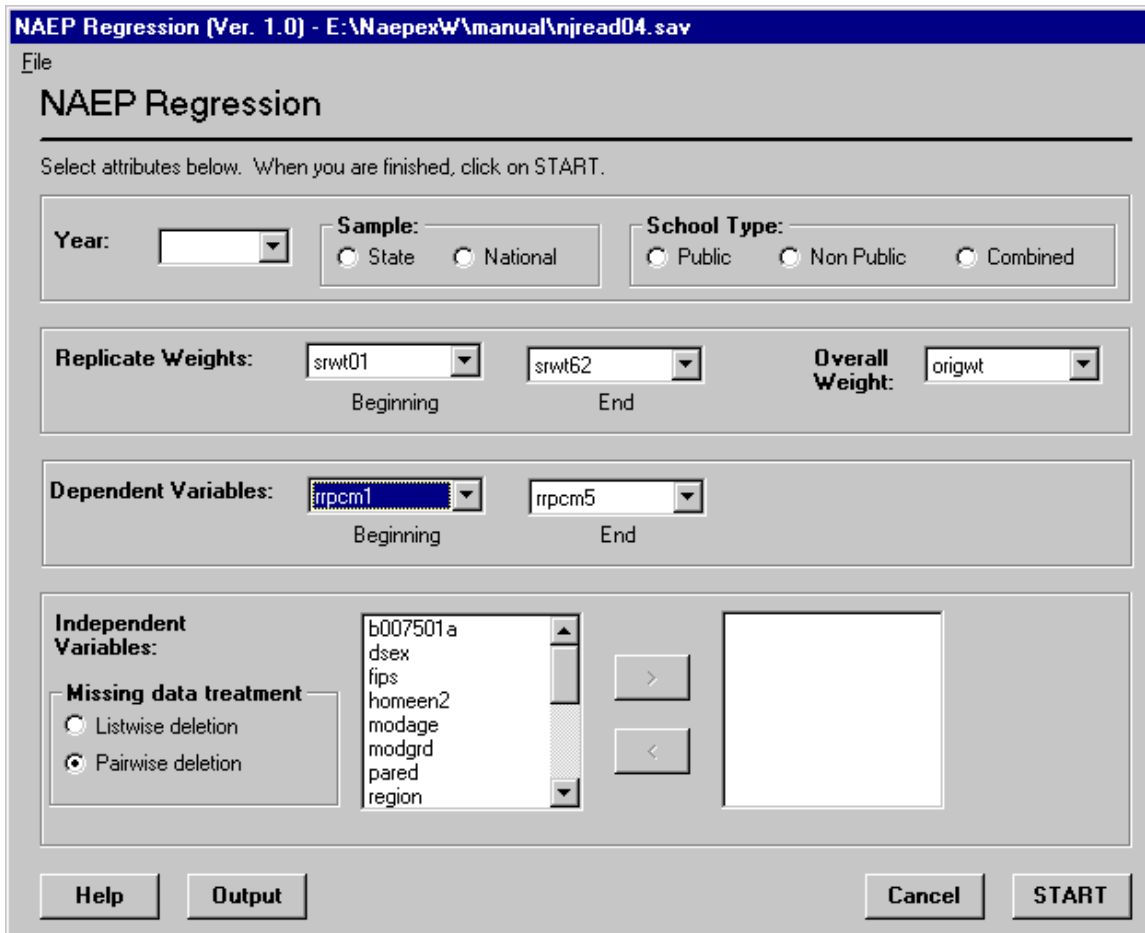


Figure 5-4: The NAEP Regression Interface

5.2.1 Sample Information

Here the user can communicate some information about the sample to the NAEP Regression Module. The first box lists the year of the sample. There are two push buttons to describe whether the sample is a state or national sample and what type of school the students who are in the sample attended (public/nonpublic/combined). The module does not use any of this information in the regression calculation. It is here so that the output can be properly labeled when the analysis is complete.

5.2.2 Weight Information

There are two types of weights necessary to the proper use of the NAEP Regression Module: replicate weights and the overall weight.

Based on the data dictionary, the NAEP Regression Module will make a good guess at what the weight values should be for a given data set and has them selected by default. The program does, however, let you change these selections by clicking on the combo box as shown in Figure 4-5. In general, you should use the default selections unless you have good reason not to. For example,

you might have created a data file with more than one set of replicate and overall weights and it is necessary to select a specific set of weights for the data you are analyzing.

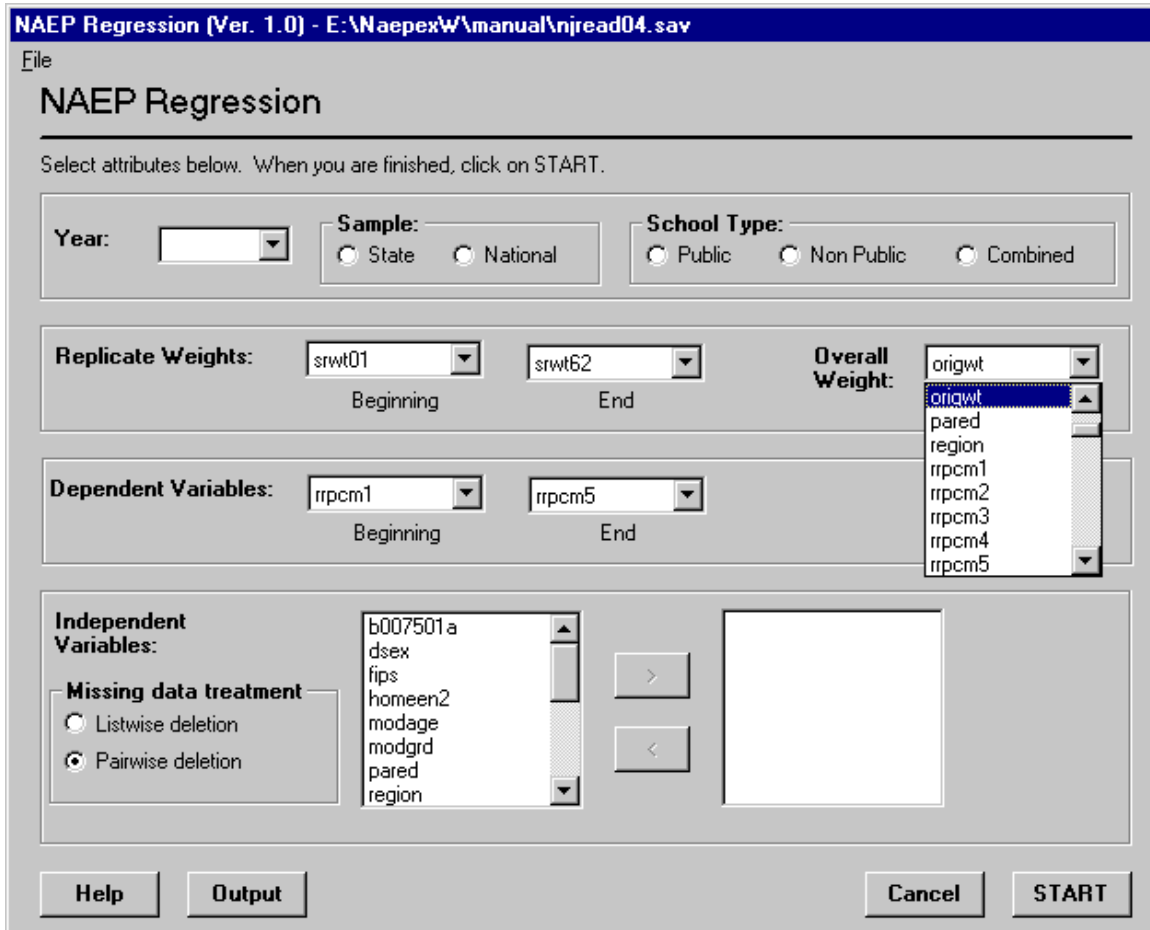


Figure 5-5: Changing the Overall Weight

5.2.3 Dependent Variables

The NAEP Regression Module assumes that the default dependent variables are the plausible values associated with the data set being analyzed. Like the weights, the module makes an attempt to figure out which NAEP IDs are the plausible values and makes them the default selection. In general it is not necessary to change these selections, but it could be possible, if the sample you are using contains more than one set of plausible values. This might happen if scores are being reported for more than one scale for a given subject, and the plausible values from more than one scale were included as part of the selection when the data set was created.

5.2.4 Independent Variables

The NAEP Regression Module lists all of the available variables in the data set in a list box on the left-hand side. To add a NAEP ID to the analysis, select it from the box on the left-hand side and then click the arrow to move it over to the box on the right-hand side. Similarly, a variable

that is selected on the right-hand side can be moved back to the left-hand side by clicking on the appropriate arrow. There must be at least one variable selected to perform the analysis.

The NAEP Regression Module supports both pair-wise and list-wise deletion. Pair-wise deletion includes in the regression calculation those records that had one or more missing independent variables. List-wise deletion throws the entire record away if any of the independent data are missing. Which records were kept and which ignored are displayed after the calculations are complete (see Section 5.2). It should be noted that a record is thrown out if any of its dependent variables or any of its weight variables are missing. See Section 5.3 for a more complete discussion of missing values.

If more than one variable is included in the analysis, and there is an explicit relationship between that variable and another variable, or one of the variables in the analysis is constant, the module will print an error stating that "it was unable to sweep variable x" where x is a number from one to the number of variables included in the analysis. The offending variable must be removed before the analysis can proceed.

5.3 Running the Analysis and Getting Results

When all of the data have been properly entered, a NAEP regression can be calculated. To start the process, simply click the Start button (**not** the Windows Start button) in the lower right-hand corner of the module screen. Once the analysis begins you are kept aware of its progress and can cancel it at anytime by pressing the Cancel button.

If the NAEP Regression Module was launched from SPSS, the process will begin immediately. If, however, the module was started in stand-alone mode, a prompt is displayed asking for a name for the output file (Figure 5-6). A default file extension of .TXT will be assigned to the file, which will be saved as a standard text file that is readable by any editor.

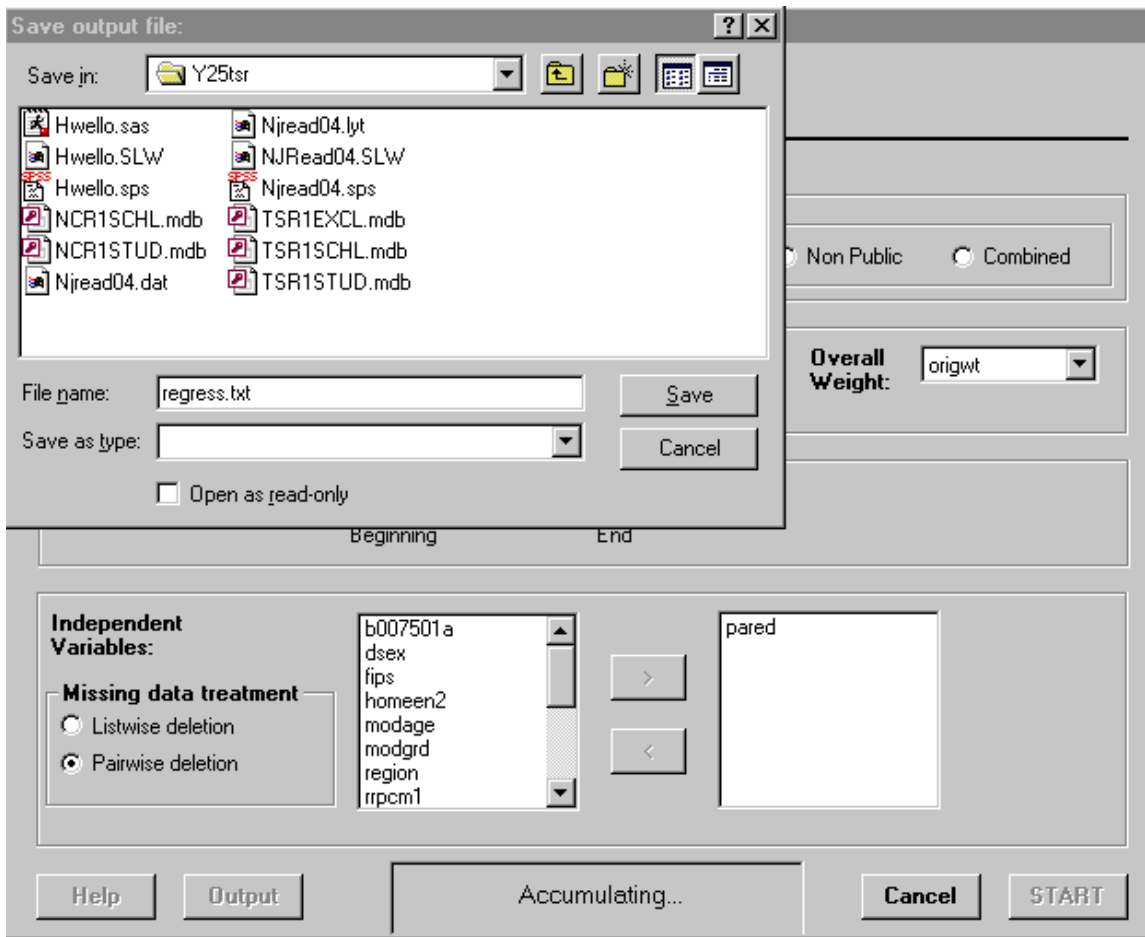


Figure 5-6: Saving the Output in Stand-Alone Mode

If the NAEP Regression Module was launched from the NAEP menu in SPSS, the output is sent immediately to an SPSS Output window (Figure 5-7). The module is then immediately unloaded.

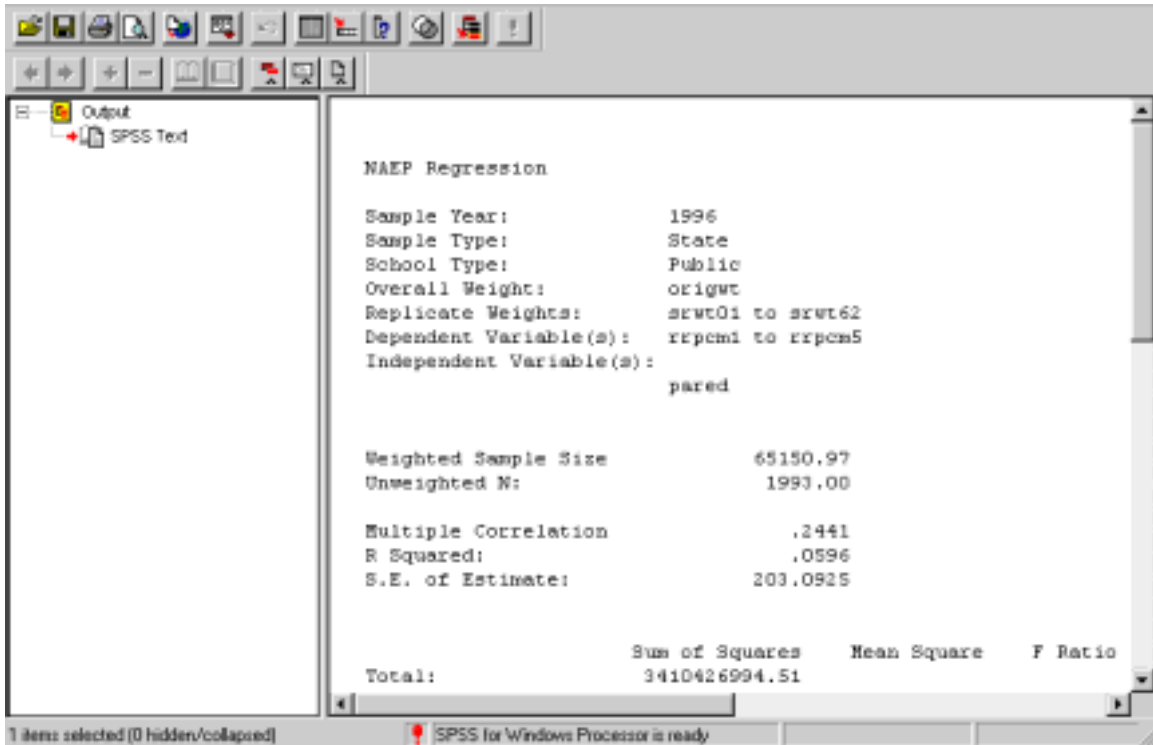


Figure 5-7: Results Displayed in SPSS Output Window

If the NAEP Regression Module was launched in stand-alone mode, the output is sent to the output file and a copy of the file is displayed in the module (Figure 5-8). Pressing OK will close the output box and then pressing cancel will close the NAEP Regression program.

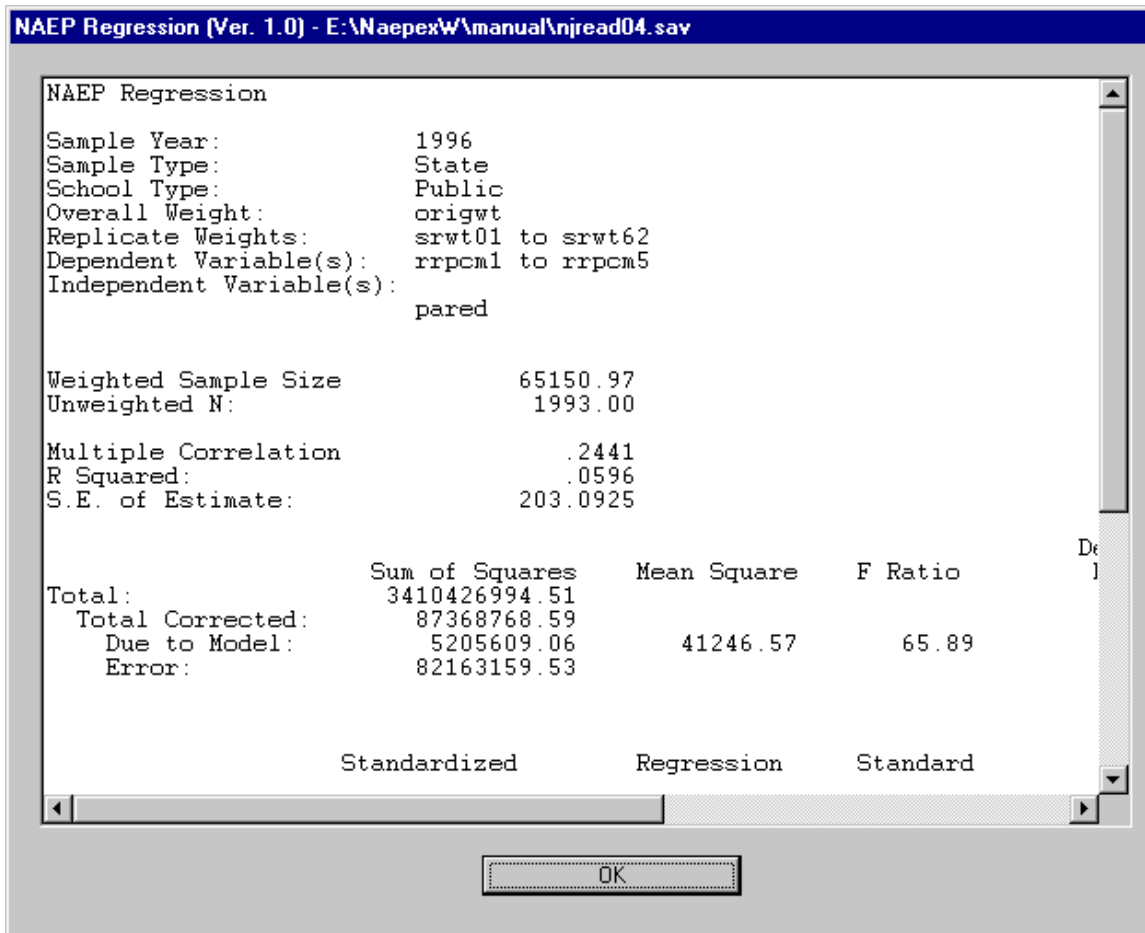


Figure 5-8: Results Displayed in Stand-Alone Module Window

5.4 NAEP Regression Tips

5.4.1 Selecting Independent Variables

The selection of appropriate independent variables is very important. A regression that includes the raw DRACE variable (race/ethnicity) would not make any sense, for example, because the assignment of arbitrary numbers to various ethnicities is essentially random and one number has no implicit relationship to the other or to the race they represent. A variable like PARED (parents' education level) is a better choice for regression analysis because as the response values increase so does the level of the parents' education (in general).

The NAEP Regression Module does not support immediate contrasting of variables. Any contrasting you might want to perform must be done from within the data analysis package itself (e.g., SPSS) by adding your own variables and applying your own transformations.

5.4.2 "Missing" Data

There are two types of "missing" data:

- There are no valid data at all (SYSMISS).
- There are data but the data itself should be considered missing. This is common in NAEP. For example a tester might circle two answers in response to a question. This data is coded as "Multiple Response", which for all practical purposes, means that data point is missing, but it has an actual code in the data set.

It is the analyst's responsibility to handle "missing" appropriately. The only data the NAEP Regression Module currently considers "missing" is the SYSMISS value of SPSS. That is, if any given variable has codes that stand for essentially missing data (e.g., MULTIPLE RESPONSE) those data will be included in the regression as if they are real data points. Therefore, the analyst should transform any data he or she considers missing to the SPSS SYSMISS value.

The NAEP Regression Module does not accumulate any record that has a SYSMISS value for any weight or plausible value.

Appendix: Statistical Methodologies

Special properties of the NAEP design affect the validity of conventional statistical inference. There are two reasons standard statistical procedures should not be applied to NAEP data. First, a complex sampling scheme, rather than simple random sampling, was used to collect the data. Second, because scaling models were used to summarize performance, measurement error must be taken into account when analyzing scale-score values.

In the NAEP sampling scheme, students do not have an equal probability of being selected. Therefore, as in all complex surveys, each student has been assigned a sampling weight. When computing descriptive statistics or conducting inferential procedures, one should weight the data for each student. Performance of statistical analyses without weights can lead to misleading results.

NAEP uses a multi-stage sampling scheme that results in observations not being independent from one another. Therefore the use of standard formulas for estimating standard errors will result in values that are generally too small.

The following provides detailed information about the methodology used to calculate the crosstabulation and regression statistics in the NAEP Data Toolkit modules. More information is available in the technical reports for the NAEP assessments.

NAEP Crosstabulation Module Methodology

The t -statistic used by the crosstabulation module in significance testing is defined by:

$$t_{df} = \frac{(A - B)}{\sqrt{S_A^2 + S_B^2}}$$

where A and B are statistics being compared and S_A and S_B are the respective jackknife standard errors. The degrees of freedom of this t -test are defined by a Satterthwaite approximation as follows:

$$df = \frac{\left(\sum_{K=1}^2 S_K^2 \right)^2}{\sum_{K=1}^2 \frac{S_K^4}{df_K}}$$

with df_K defined as follows:

$$df_K = F_K \left(\frac{\left(\sum_{j=1}^n (t_j - t)^2 \right)^2}{\sum_{j=1}^n (t_j - t)^4} \right)$$

where

$$F_K = \left(3.16 - \frac{2.77}{\sqrt{M}} \right)$$

In the above equations, t_j is the j^{th} replicate estimate for the mean of a subgroup, and t is the estimate of the subgroup mean using the overall weight and first plausible value and n is the number of replicate weights or PSU pairs. M is the number on non-empty PSU pairs. So M is less than or equal to n , the total number of PSU pairs. F_K is a factor that adjusts for the downward bias of the degrees of freedom estimate.

The False Discovery Rate (FDR) Multiple Comparison Procedure

When many different groups are compared (i.e., multiple sets of confidence intervals are being analyzed), statistical theory indicates that certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., .05), adjustments-called “multiple comparison procedures”-must be made. One such procedure, the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995)¹ was used to control the certainty level.

Unlike the other multiple comparison procedures (e.g., the Bonferroni procedure) that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, familywise procedures are considered conservative for

¹ Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1), 289-300.

large families of comparisons (Williams, Jones, & Tukey, 1999).² Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures.

The 1998 assessment is the first time NAEP has used the Benjamini-Hochberg procedure to maintain FDR for all multiple comparisons. Prior to the 1996 assessment, the Bonferroni procedure was used for multiple comparisons. In 1996, either the Bonferroni or Benjamini-Hochberg FDR procedure was used, depending on the testing situation. The Benjamini-Hochberg FDR procedure was used for large numbers of comparisons (i.e., any comparisons involving all of the states): (a) all pairwise comparisons of the states; (b) all comparisons of individual states to the national average; and (c) the trend for each state, which compared the current mean for the state to the state's mean in the previous assessment. All other multiple comparisons for the 1996 assessment used the Bonferroni procedure. The 1994 NAEP reading assessments used the Bonferroni procedure exclusively for multiple comparisons.

The Benjamini and Hochberg application of the FDR criterion can be described as follows. Let q be the number of significance tests made and let $P(1) \leq P(2) \leq \dots \leq P(q)$ be the ordered significance levels of the q tests, from lowest to highest probability. Let α be the combined significance level desired, usually .05 for one-tailed tests (or .025 for two-tailed tests). The procedure compares $P(q)$ with α , $P(q-1)$ with $\alpha (q-1)/q$, . . . , $P(j)$ with α_j/q , stopping the comparisons with the first j such that $P(j) \leq \alpha_j/q$. All tests associated with $P(1)$, . . . , $P(j)$ are declared significant; all tests associated with $P(j+1)$, . . . , P_q are declared nonsignificant.

NAEP Regression Module Methodology

NAEP uses a multistage cluster sampling scheme which makes traditional methods of variance calculation inappropriate. The method of variance estimation used for the NAEP design is a jackknife procedure.

The computation of the jackknife procedure involves replicate weights. For notation, we will refer to the replicate weights as w_{j1} to w_{jm} , where m is the number of PSU pairs and j is the j^{th} student. The overall weight will be w_{j0} in this notation.

We also have a series of plausible values as different estimates of a student's ability for a certain subject area (for example, MRPCM1 to MRPCM5 for the mathematics composite score). We will use the notation y_{j1} to $y_{j,nimp}$, where $nimp$ is the number of different ability estimates (5 in 1992) and j is for the j^{th} student. We will let x_j represent a vector

² Williams, V. S. L., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, 24(1), 42-69.

of observed variables for the j^{th} person, to be used as independent variables in the regressions.

At the present time the regression module will use the plausible values as the dependent values only. If a plausible value were used as an independent variable, that coefficient would have to be adjusted to account for error in the independent variable (see Draper & Smith, 1981, pp. 122-124)³.

The regression coefficients estimated are the average of the coefficients of the five weighted regressions where each regression uses w_{j0} and a different plausible value as the dependent variable. So if we let W_0 be the diagonal matrix of weights w_{j0} , X the matrix of independent variable vectors x_j , and Y_i the vector of the i^{th} plausible value for each person in the regression, then

$$b_{0i} = (X'W_0X)^{-1} X'W_0Y_i.$$

So b_{0i} is a vector of regression coefficients using the weights matrix W_0 and the plausible value vector Y_i . The average of the *nimp* b_{0i} s is \bar{b}_0 :

$$\bar{b}_0 = \frac{\sum_{i=1}^{nimp} b_{0i}}{nimp}$$

To calculate the standard errors of the coefficients and the F -test, we need to do the following. Calculate b_{k1} , a vector of regression coefficients using plausible value 1 and replicate weight k , in the formula

$$b_{k1} = (X'W_kX)^{-1} X'W_kY_1$$

where W_k is the diagonal matrix of replicate k weights.

We then form the vector of differences $(b_{k1} - b_{01})$. The variance-covariance matrix, due to sampling error, of the regression coefficients can be computed by:

$$V_{samp} = \sum_{k=1}^m (b_{k1} - b_{01})(b_{k1} - b_{01})'$$

³ Draper, N. R., & Smith, H. (1981). *Applied regression analysis*. New York: John Wiley & Sons.

This is a $(p+1) \times (p+1)$ variance-covariance matrix due to sampling error, where p is the number of independent variables. Then we can compute V_{imp} , the $(p+1) \times (p+1)$ variance-covariance matrix

$$V_{imp} = \frac{\sum_{i=1}^{nimp} (b_{0i} - \bar{b}_0)(b_{0i} - \bar{b}_0)'}{nimp - 1}$$

and the total variance-covariance matrix

$$V_{tot} = V_{samp} + \frac{nimp + 1}{nimp} V_{imp}.$$

So the corresponding estimates of the standard errors of \bar{b}_0 are the elements of the vector formed by the square root of the diagonal elements of V_{tot} .

$$\widehat{SE}_{\bar{b}_{0c}} = \sqrt{V_{tot_{cc}}}.$$

To do a t -test with m (the number of PSU pairs, or replicate weights) degrees of freedom, you would compute

$$t_c = \frac{\bar{b}_{0c}}{\widehat{SE}_{\bar{b}_{0c}}}$$

where $c = 1, \dots, p+1$.

The following Wald statistic is the test for the overall fit of the regression model

$$F = \bar{b}_0^{*'} (V_{tot}^*)^{-1} \bar{b}_0^*$$

with degrees of freedom described in Section 8.4.2 of the *Technical Report of the NAEP 1992 Trial State Assessment in Mathematics*. The numerator degrees of freedom is p , and the denominator degrees of freedom is given as

$$\nu = \frac{1}{\frac{f_{nimp}^2}{nimp - 1} + \frac{(1 - f_{nimp})^2}{m}},$$

where

$$f_{nimp} = (1 + nimp^{-1}) \frac{TRACE(V_{imp}^* V_{tot}^{*-1})}{p}$$

and where V_{imp}^* is the same matrix as V_{imp} with the row and column for the intercept removed, V_{tot}^* is the same matrix as V_{tot} with the row and column for the intercept removed, and \bar{b}_0^* is the same vector as \bar{b}_0 with the estimate of the intercept removed.