# SPSS for Microsoft Windows

## For New Users of SPSS on Microsoft Windows at NYU

Third Edition — September 15, 1998

Contents



ACF NYU

**Academic Computing Facility**
**New York University**
http://www.nyu.edu/acf/ on the World-Wide Web

**Publication History**
Third Edition, September 1998; Revised to reflect edition 8.0 changes, and to add Cronbach alpha, filtering for subsets of data. Revised by Frank LoPresti, Deborah Padgett and Melissa Whitney.

Second Edition, September 1997; Revised to reflect edition 7.5 of SPSS. Revised by Melissa Whitney, Alexandra Cincotta, Frank LoPresti and Deborah Padgett.
First Edition, October 1995; minor revisions, September 1996.
Written by Mary Ann Jones, Frank LoPresti, Zvia Naphtali, and Melissa Whitney; edited by Melissa Whitney, with Estelle Hochberg

# 1. Introduction

Statistical Package for the Social Sciences (SPSS) is a comprehensive integrated software package for statistical data analysis. SPSS for Windows allows you to store data, perform transformations and analyses, and produce charts and graphs of results. Data are entered using a spreadsheet and results are displayed in a separate output window. The data and the output can be saved independently for the next work session. The output tables can be copied to a word processing application for inclusion in papers.

This document is an introductory guide to SPSS version 8.0 running in the Microsoft Windows environment. It was written particularly for students in the course Social Work Research II (S44.2008). Accordingly, examples are given with this course's requirements in mind. Nonetheless, the guide will be helpful to other SPSS users as well.

This guide will be most useful if you are sitting at a computer, following the steps for working with SPSS as you read. The illustrations are taken directly from the computer screen and will match what you see on your monitor while you are working.

Classes in SPSS and Windows basics are offered by NYU's Academic Computing Facility (ACF). Schedules are available at any ACF computer lab; at the ACF HelpCenter on the second floor of Warren Weaver Hall; and via the World Wide Web at **http://www.nyu.edu/acf/classes/**.

## Windows Online Tutorial

A basic familiarity with the Windows operating system is assumed in this document. If you are new to Windows, run the online tutorial before proceeding with SPSS. To start the tutorial on any Windows machine:

1. Select **Help** from the **Start** menu.
2. Select the **Contents** tab in the Help Topics: Windows Help dialog box.
3. Select **Tour: Ten minutes to using Windows** and click the **Display** button.

## Getting Started at an ACF Lab

SPSS for Windows is available at the following ACF computer labs:

- Tisch Hall — 40 W. 4th Street, Rooms LC-7 and LC-8
- 14 Washington Place — 14 Washington Place, basement

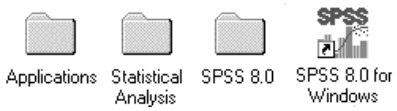
For information on eligibility to use the ACF labs, please refer to the brochure, *Academic Computing Facility: Computer Services and Internet Resources for the NYU Community*. Lab hours are listed on the last page of this guide.

Bring two 3.5 inch double-sided high density floppy disks formatted for DOS/Windows on which to save your work.

# 2. SPSS Basics

## Starting SPSS

To start SPSS for Windows at an ACF lab:

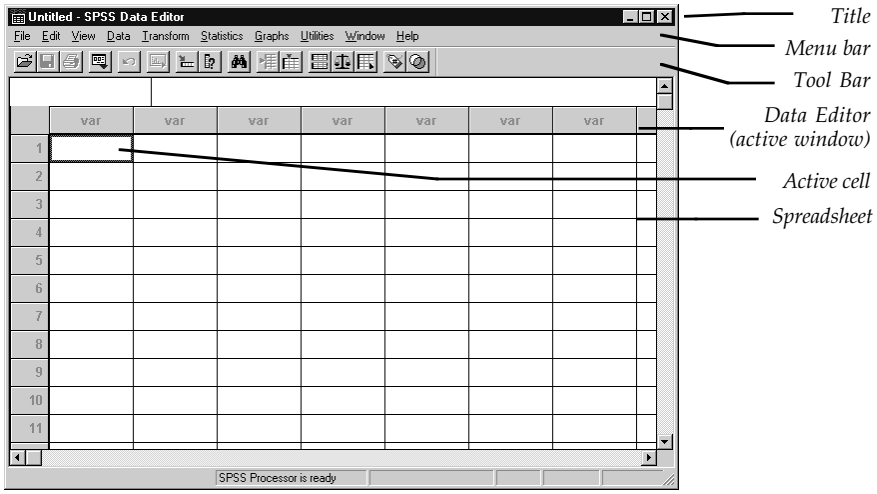



1. Double-click on the Applications folder on the desktop.
2. Double-click on the Statistical Analysis folder.
3. Double-click on the SPSS 8.0 folder.
4. Double-click on the SPSS 8.0 program icon.

The SPSS window opens.

At this point you have two options: begin creating a new data set (see page 3) or open an existing data set (see page 8). Before covering either topic we will review the elements of the SPSS Data Editor window.

*The Main Window*




SPSS opens multiple **windows** as it performs operations. The two windows you will use most frequently are the Data Editor and the Output Viewer. Every window has a title bar and a menu bar.

The **title bar** displays the name of the application, SPSS, and the current window, for example, Data Editor or Viewer. On either end of the title bar are common elements of Windows applications: control menu, minimize and maximize buttons. If these terms are unfamiliar to you, consider running the Windows tutorial (see page 1).

The **menu bar** appears just below the title bar. Clicking once on a word in the menu bar opens the associated menu, from which you can select a command. The instruction "select Save from the File menu" tells you to click once on the word "File," then once on the word "Save."

The **Data Editor** is a spreadsheet in which you define your variables and enter values. All of your data entry is done in this window. The title bar displays the name of the open data file, or "Untitled" if the file has not yet been saved.

The **active window** is the window you are currently using or that is currently selected. Keystrokes and commands are applied to the active window. If a window is active, its title bar changes color to differentiate it from other windows. Only one window is active at a time. To make a window active, click on any part of the window, or select it from the Window menu.

The **active cell** in the spreadsheet is indicated by a dark border. This is the cell in which typing will appear. Only one cell is active at a time. You can make a different cell active by clicking on it, or by using the tab, return, and cursor keys to move to it.

The **spreadsheet** which appears within the data editor is a series of rows and columns. Each column corresponds to a variable. Each row corresponds to a case or subject.

## Creating a New Data Set

Creating a new data set consists of the following steps:

1. Define variables.
   - Name.
   - Description.
2. Save the data set.
3. Produce a data dictionary (optional).
4. Enter data.
5. Save the data set again.
6. Print the cases (optional).

### Defining Variables

Variables are defined one at a time using the Define Variable dialog box.

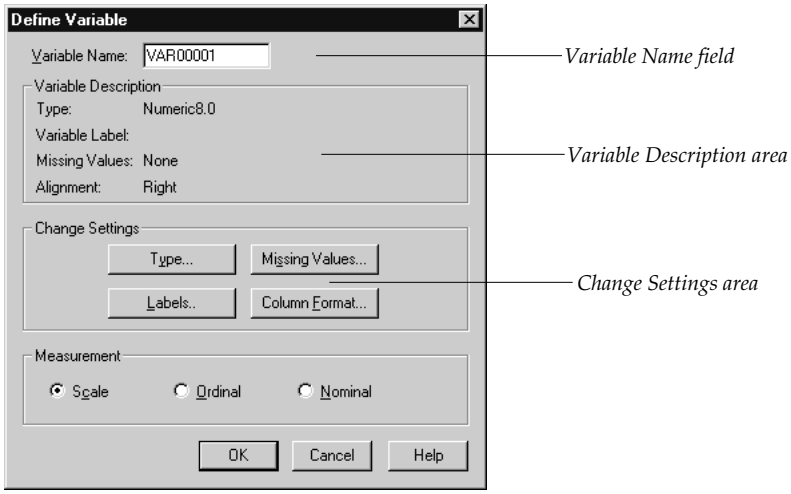
To access the Define Variable dialog box:

Double-click on the top of a column, where the word var appears.

*or*

Select Define Variable from the Data menu.

*The Define Variable Dialog Box*




Variable Name

The **Variable Name** field displays the name of the variable. You can edit the name directly by placing your cursor in this field. SPSS provides a default name for each variable in the format VAR00001, VAR00002 and so on. Replace this name with one which is more meaningful to your research. The variable name must begin with a letter of the alphabet and cannot exceed 8 characters. Spaces are not allowed within the variable name. Each variable name must be unique. Make your variable names memorable. SALSTART and SALNOW are easier to interpret than SALARY1 and SALARY2.

Variable Description

The **Variable Description** area displays information about the variable. The default variable description is a Numeric variable with a column width of 8, right justified; no decimal places; no labels; and no missing values. This description will suffice for the majority of your variables. To modify the variable description, use the buttons in the **Change Settings** area.

The aspect of the variable description that you most frequently will change is Labels. There are two types of labels in SPSS: Variable and Value.

The Variable Label can be thought of as a long name for the variable. The eight character limit on the variable name can result in arcane names such as PAEDUC or AGCHLD2. The variable label shows that PAEDUC stands for "Father's Education Level" and AGCHLD2 stands for "Age of the Second Child." Variable labels can be up to 120 characters long and can include spaces (which variable names cannot). If you enter a variable label, the label will print on your charts and reports instead of the name, making them easier to understand.

Value Labels provide a key for translating numeric data that is not inherently meaningful. For example, a variable representing age contains numeric values indicating age in years. These numbers are easily interpreted. Another variable indicates a subject's sex. A value of 1 does not

automatically reveal whether the subject is male or female, so value labels must be supplied.

The other elements of the variable description are type, missing values, and alignment. These options are discussed in the online SPSS help system. To access the help system, from the Help menu, select Topics then Contents.

## Example: Creating a Sample Data Set

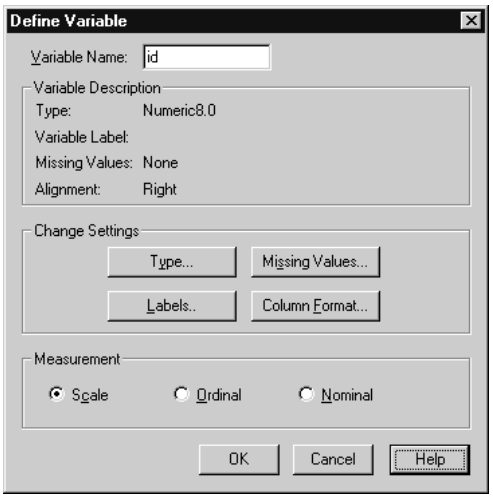
To introduce you to variable definition, we'll create a simple data set using information on age, gender and salary.

**Define a variable named id, which will serve as a counter for our cases.**

1. Double-click on the top of the first column, where the word var appears.

   *or*

   Click anywhere in the first column, then select Define Variable from the Data menu. The Define Variable dialog box opens.




2. Enter **id** in the Variable Name field. The default description is suitable for the id variable, so we will close the dialog box.
3. Click on the OK button. You are returned to the data window, and the first column now contains the variable id.

**Define a variable named age.**

1. Double-click on the top of the second column, where the word var appears.

   *or*

   Click anywhere in the second column, then select Define Variable from the Data menu. The Define Variable dialog box opens.

2. Enter **age** in the Variable Name field.
3. Click on Type, then set the width as 8.
4. Click on the OK button. You are returned to the data window, and the second column contains the age variable.

**Define a variable named gender as a numeric variable with value labels.**

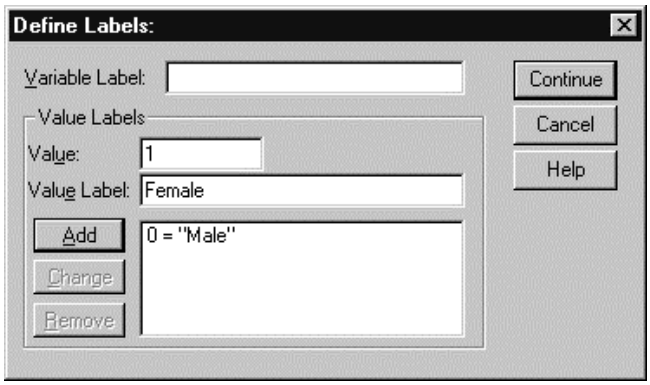
1. Double-click on the top of the third column, where the word var appears.

   *or*

   Click anywhere in the third column, then select Define Variable from the Data menu. The Define Variable dialog box opens.

2. Enter **gender** in the Variable Name field.

3. Click on the Labels button in the Change Settings area. The Define Labels dialog box opens.




4. The variable name "gender" is sufficiently descriptive, so we will not add a Variable Label. Leave the Variable Label box empty.

5. Enter **0** in the Value field and **Male** in the Value Label field. Click the Add button. The phrase 0="Male" appears in the lower section of the dialog box. Now, enter **1** in the Value field and **Female** in the Value Label field. Click on the Add button again.

   We are using the values 0 and 1 to represent the two genders, but the choice of numbers is arbitrary. Numeric values are often assigned to a categorical variable at the time a questionnaire is designed. When that is the case, use the preassigned numbers to define value labels.

6. Click on the Continue button. The Define Labels dialog box closes and you are returned to the Define Variable dialog box.

   NOTE: Do not to click the Continue button until all of your labels appear in the lower section of the dialog box. If you do not click on the Add button after entering a Value and a Label, the information will be lost.

7. Click on the OK button. You are returned to the data window, and the third column now contains the gender variable.

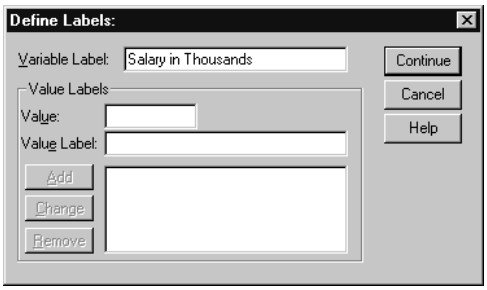
**Define a variable named salary with a Variable Label "Salary in Thousands."**

1. Double-click on the top of the fourth column, where the word var appears.

*or*

Click anywhere in the fourth column, then select Define Variable from the Data menu. The Define Variable dialog box opens.

2. Enter **salary** in the Variable Name field.
3. Click on the Labels button. The Define Labels dialog box opens.




4. Enter **Salary in Thousands** in the Variable Label field.
5. Click on the Continue button. The Define Labels dialog box closes and you are returned to the Define Variable dialog box. Notice that the variable label you just entered appears in the variable description area of the Define Variable dialog box.
6. Click on the OK button. You are returned to the data window, and the fourth column now contains the salary variable.

All of the variables in our sample data set are now defined. The remaining steps to complete the data set are:

1. Save the data set.
2. Produce a data dictionary (optional).
3. Enter data.
4. Save the data set again.
5. Print the cases (optional).

## Saving the Data Set

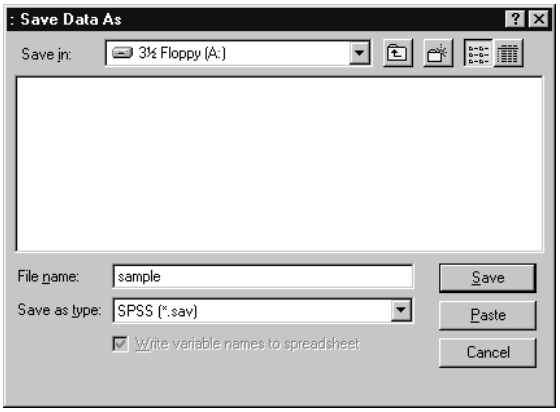
Work performed on your data set lasts only during the current session. To retain your work you must save it to a file. It is a good idea to save your work frequently. Machines do occasionally go down, and any work not saved will have to be redone.

To save a new data set:

1. Insert your floppy disk in the A:\ drive.
2. Make sure that the data editor is the active window. The active window appears on top of other windows and has a differently colored title bar. If the data editor is not active, select it from the Window menu.
3. Select Save from the File menu. The Save Data As dialog box opens.
4. From the Save as Type drop-down list, select **SPSS (*.sav)**.

From the Save in drop-down list, select **a:** for the floppy drive.

In the File name box, enter a name for the file. SPSS automatically adds the extension .sav. For our sample data set, enter the name **sample**.



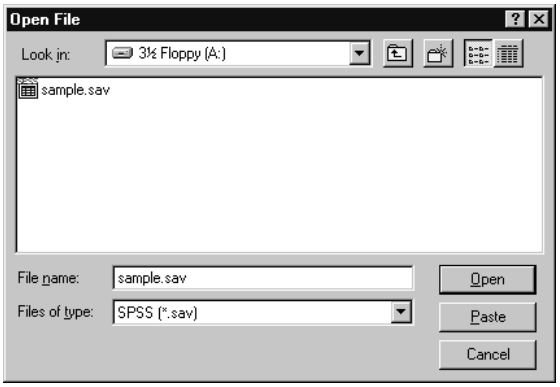

5. Click Save. It will take a few moments for the computer to save the file to disk.

If you have previously saved the file you are working on, selecting Save from the File menu automatically saves your changes to the appropriate file. If you want to save your changes to a different file, use the Save As command on the File menu.

## Opening a Saved Data File

To open a data file which has been saved to a floppy disk:

1. Insert your floppy disk in the A:\ drive.
2. Select Open from the File menu. The Open File dialog box opens.




3. From the File of type drop-down list, select **SPSS (*.sav)**

   From the Look in drop-down list, select **a:**

   In the File name box, enter the name of the file to open. You do not need to include the ".sav" extension.

4. Click Open. It will take a few moments for the computer to open the file from the disk.

## Producing a Data Dictionary

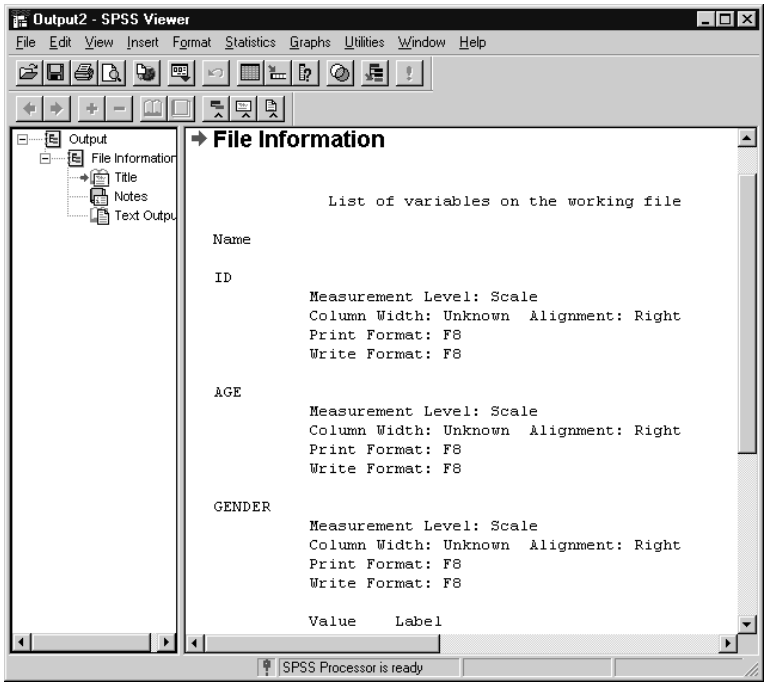
A data dictionary is a list of the parameters for each variable in your data set: the variable name, type, variable label, value labels, missing value definition, and display format. The data dictionary documents how each variable in your data set was defined.

Producing a data dictionary at this point allows you to quickly check each of your variable definitions for errors before inputting data.

To produce a data dictionary:

- Select File Info from the Utilities menu.

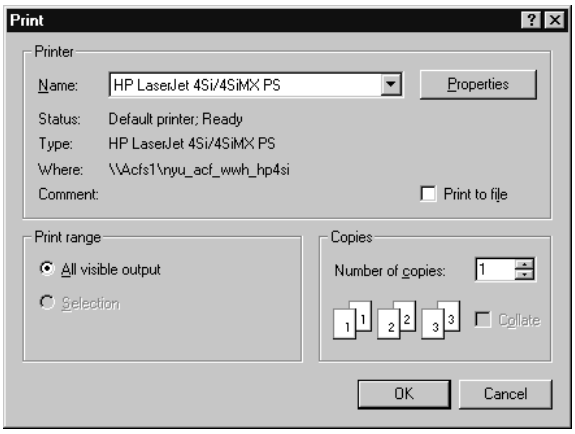
SPSS opens the output viewer, bringing it to the foreground and covering the data editor. The data dictionary begins with the line "List of variables on the working file." To the left is the name of each variable, and to the right the column number in which the variable appears. Below each variable name is the print and write format, followed by any special characteristics of the variable description, for example, value labels. Use the scroll bars to view the entire data dictionary.




## Printing the Data Dictionary

To print the data dictionary:

1. Make sure that the output viewer is the active window. If it is not, select the output viewer from the Window menu.
2. Select Print from the File menu. The Print dialog box opens. Before printing the data dictionary, check which printer your machine is connected to by looking at the Where line in the Print dialog box. If you are still uncertain about where material will print, ask an ACF Lab consultant for help.




3. Click OK.

The entire contents of the output window print out. To print only the most recent output (in this case, the data dictionary):

1. Select the Select command on the Edit menu, then select Last Output on the Select submenu. The last output is highlighted in the output window.
2. Select Print from the File menu. In the Print dialog box select **Selection** before clicking OK. Only the highlighted text prints out.

To return to the data editor, select the data editor from the Window menu.

## Entering Data

Once all of the variables are defined, you can begin entering data. The data is typed into the spreadsheet one cell at a time. Each cell represents one respondent's answer to one question. A column represents a single piece of information about every respondent, and a row represents every piece of information about a single respondent.

When you type information into the data window of SPSS for Windows it appears in the edit area at the top of the window. The information is entered into the cell when you move the active cell. The mouse and the tab, return, and cursor keys can be used to enter data.

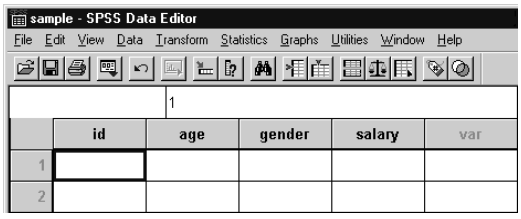
| | |
|---|---|
| Mouse | activates the cell clicked on |
| Tab key | moves the active cell one cell to the right |
| Return key | moves the active cell one cell down |
| Cursor keys | move the active cell one cell in the direction of the arrow. |

## Example: Entering Data into the Sample Data Set

To introduce you to how to enter data in SPSS, we'll enter some values for our sample data set.

**Enter values for the first case.**

1. Click on the first cell to make it the active cell.
2. Type the number **1**. Notice that it appears in the edit area at the top of the screen.

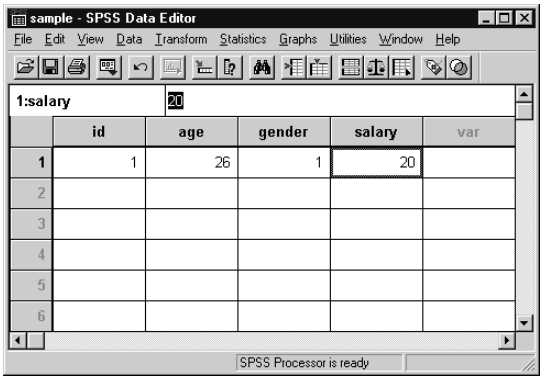



3. Press the Tab key. The number 1 appears in the first cell, and the active cell moves right one column.

   The row number to the left of the first column turns black. This indicates that a case has been created.

   The second cell in the first row is now the active cell. Notice that in the box to the left of the edit area the caption "1:age" is displayed. This indicates that the active cell is in the first row and the age column.

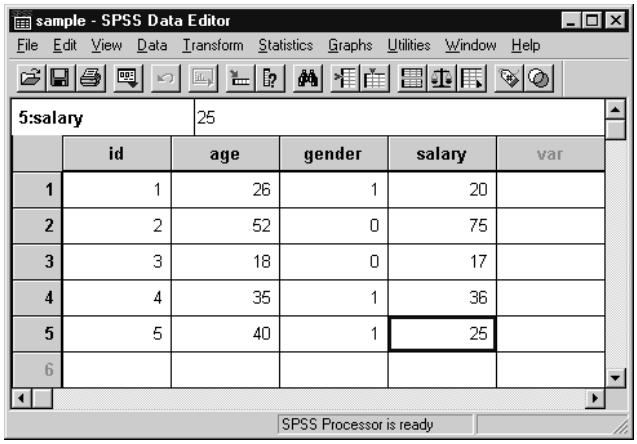
4. Type the number **26**. This information appears in the edit area.
5. Press the Tab key. The number 26 appears in the second cell, and the active cell moves one cell to the right.
6. Type **1** for the cell "1:gender" and press the Tab key.
7. Type **20** for the cell "1:salary" and press the Tab key.




We now have a complete record for a 26-year-old female making $20,000.

**Complete the data set.**

Enter values for the next four cases according to the spreadsheet below:




Experiment with the tab, return and cursor keys and with the mouse to find the method most efficient for your work.

### Saving the Data Set Again

Once you have entered several cases it is a good idea to save your work again. Select Save from the File menu. Because you have already named and saved this file, the Save dialog box does not open. The new information is automatically saved to the correct file and disk.

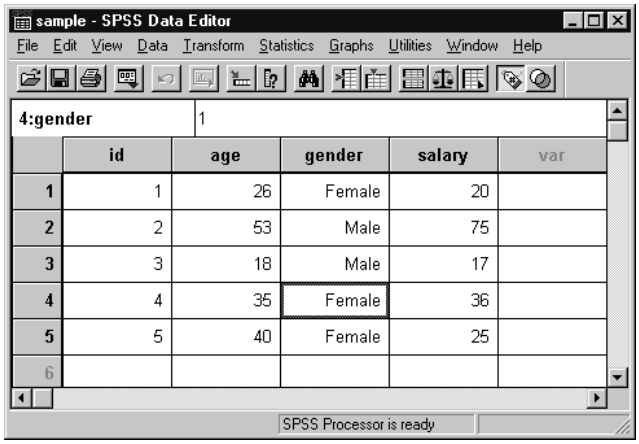
## Displaying Value Labels

To display value labels for the data in the spreadsheet:

Select Value Labels from the View menu.

*or*

Click on the Value Labels button on the toolbar. 

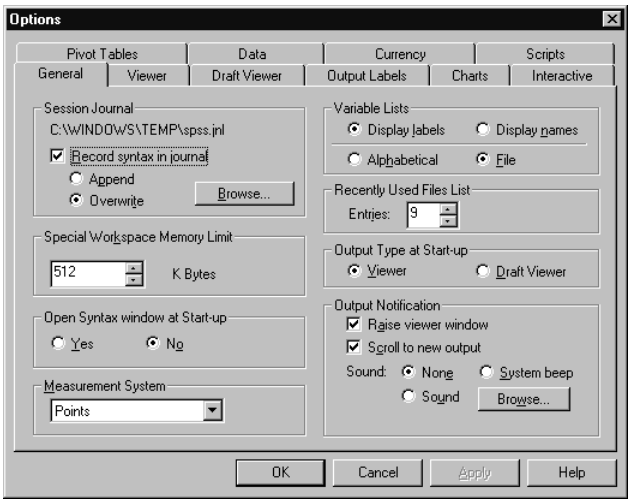



The gender column displays the text value labels for the numeric data entered earlier. In the illustration above, the active cell is on "Female," but notice that the underlying numeric value (1) displays in the edit area on the top left of the window.

## Changing the Order of Variables

For most of the statistical analyses in SPSS, you select variables to include from a list in a dialog box. This list may appear either in alphabetical order or in the order in which the variables occur in the original data file. If you have hundreds of variables, you would probably choose to view the list in alphabetical order, while if your data set is well known to you, the data entry order of the variables may be more useful. To change the order in which SPSS displays variable names:

1. Select Options from the Edit menu. The Options dialog box opens.




2. In the Variable Lists area select either the **Alphabetical** or the **File** option and click **OK**.

   SPSS displays a message that the change will take effect the next time you open a data file.

3. Close the current data file by selecting New from the File menu, then Data from the New sub menu, or by closing and re-opening the SPSS program.

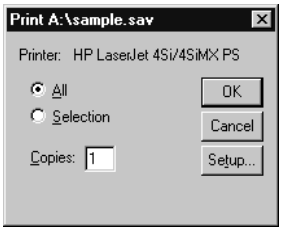
4. Re-open your data file by selecting Open from the File menu.

## Printing the Spreadsheet of Cases

To print the complete spreadsheet of cases you have entered:

1. Make sure the data editor is the active window. If it is not, select the data editor from the Window menu to activate it.

2. Select Print from the File menu. The Print dialog box opens. The title bar of the dialog box displays the name of the window to print.




3. Click OK.

The entire spreadsheet prints out. To print only certain rows or columns, use the mouse to select them, then check the Selection option in the Print dialog box before clicking OK.

## Common Errors

**Empty cases at the bottom of the spreadsheet**

When you enter a value anywhere in a row, SPSS automatically creates a case for that row. If you then delete the value, the cell contents are deleted, but the row is still considered a case by SPSS. It is a case consisting entirely of missing values. Any analysis run on your data set will contain erroneous missing values.

To remove an empty case, click on the row number to the left of the case. The row is highlighted. Press the delete key. The case is deleted, and the row number appears in grey instead of black.

**Unwanted variables**

When you enter a value in a column where a variable has not been defined, SPSS automatically creates a dummy variable with a name such as "VAR00001." If you then delete the value, the variable still exists.

To remove an erroneous variable, click once on the grey cell containing the variable name (VAR00001). The column is highlighted. Press the delete key. The column is deleted and the heading returns to a grey "var" placeholder.

**Inserting a variable (column) in the middle of the spreadsheet**

To insert a variable, click on the column before which you want the new variable to appear. Select Insert Variable from the Data menu. A column is inserted to the left of the active cell. Double-click on the grey header of the column to define the new variable.

**Inserting a case (row) in the middle of the spreadsheet**

To insert a case, click on the row above which you want the new case to appear. Select Insert Case from the Data menu. A row is inserted above the active cell. Enter values as you would for any other case.

# 3.Working With the Data Set

Now that you are familiar with how to enter data and get around within SPSS, you are ready to use the system to perform variable transformations and analyze data.

This section uses a shorter notation for making selections from menus. The instruction "Select Define Variable from the Data menu" is shortened to:

Select Data > Define Variable.

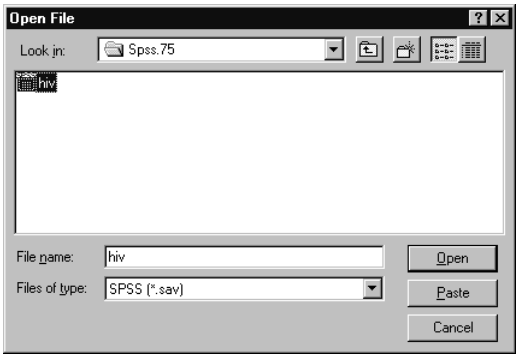
For a command on a submenu, the notation takes the form of "Menu Title > Submenu Title > Command Name."

We will be working with the hiv.sav data file in this section.

### Opening the hiv.sav file

To open the hiv.sav file on the NYU-ACF network:

1. Select File > Open. The Open File dialog box opens.




2. From the File of type drop-down list, select **SPSS (*.sav)**.

   From the Look in drop-down list, select **x:**, then **Courses**, then **S442008**.

   In the File name box, type in the name **hiv**.

3. Click Open. It will take a few moments for the computer to open the file from disk.

## Variable Transformations

### Recoding

There are three main reasons for recoding variables: to compress categorical data values into a smaller number of categories; to reverse the values of some items in a Likert-type scale preparatory to calculating a sum or mean; and to convert a continuous variable into a grouped or categorical variable.

In the first instance, you might run a chi-square analysis and find that more than 20% of the cells have expected values of less than 5. Recoding to compress some of the categories into a single category will enable you to reduce the number of cells with expected values of less than 5.

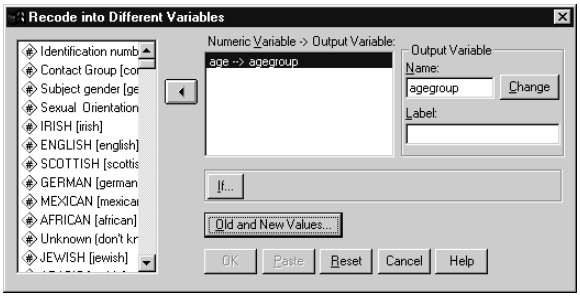
In the second instance you may have a self-esteem inventory in which all the questions are ranked on a scale of 1 to 5. However, because of the way the questions are phrased, the value 1 indicates high self-esteem on some questions and low self-esteem on others. Before summing across the questions, you would recode the scores on the items that requiring reversing, so the numerical values represent the same level of self-esteem.

Finally, you might want to present a continuous variable as a grouped or categorical variable. For example, a continuous variable containing individuals' ages might be recoded into age groups.

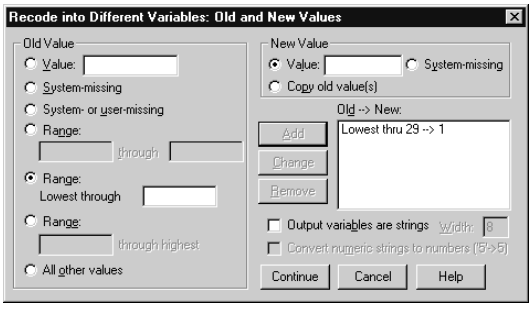
## Recoding a Variable into a Smaller Number of Categories

To recode the hiv data set variable "age" into three categories:

1. Select Transform > Recode > Into Different Variables. The Recode into Different Variables dialog box opens.




2. Select **age** from the variable list and click on the arrow button, so region appears in the Numeric Variable -> Output Variable box.
3. Enter **agegroup** in the Output Variable Name box. This variable will be created, and the recoded values calculated and entered for each case. Click on the Change button. The variable agegroup is added to the expression in the center box.
4. Click on the Old and New Values button. The Recode into Different Variables: Old and New Values dialog box opens. Here we will recode the values under 30 years of age in one group, from 30 to 50 into another and over 50 into a third.

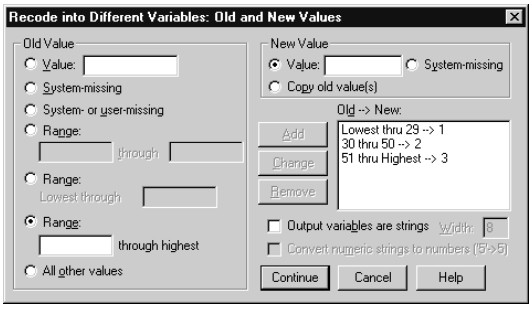



5. In the Old Value area, select Range: Lowest through and enter **29**.
6. In the New Value area, enter **1** in the Value box. The value 1 will represent under 30 in our new variable.

7. Click on the Add button. The expression "Lowest thru 29 -> 1" appears in the Old -> New area.

8. In the Old Value area, select Range: through. Enter **30** as the low value and **50** as the high value. In the New Value area, enter **2**. The value 2 will represent the group aged 30 to 50 in our new variable. Click the Add button. The expression "30 thru 50 -> 2" is added to the Old -> New area.

9. In the Old Value area, select Range: through highest and enter **51**. In the New Value area, enter **3**. Click the Add button. The expression "51 thru Highest -> 3" is added to the Old -> New area.




10. Click on the Continue button. The Old and New Values dialog box closes, and you are returned to the Recode into Different Variables dialog box.

11. Click OK. The Recode into Different Variables dialog box closes and SPSS activates the data editor. The new variable agegroup appears in the last column of the spreadsheet. Each subject younger than thirty has been given a value of 1, and each subject aged thirty to fifty, a value of 2, and each subject over fifty, a value of 3. The values 1, 2 and 3 are not inherently meaningful, so the next step is to assign value labels.

12. Double-click on the header of the agegroup column to open the Define Variable dialog box. Click on the Labels button to open the Define Labels dialog box. Enter the value labels **1="Under 30"**, **2="30 to 50"** and **3="Over fifty"**. Click the Continue button to close the Define Labels dialog box, then the OK button to close the Define Variable dialog box.

    If you have the Value Labels option on the View menu turned on, the labels will display in the spreadsheet.

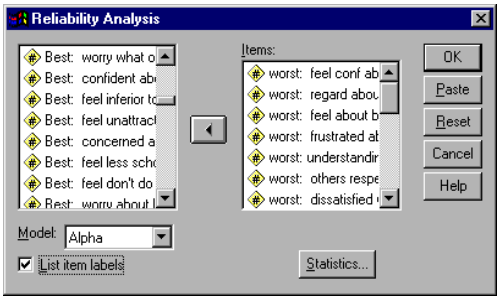
## Adding Across Variables

A major reason for adding across variables is to generate a single scale score from multiple Likert-type items. For example, in a self-esteem inventory, the answers to multiple questions are summed to arrive at a total self-esteem score. You might also use the mean of the answers as the score. Taking the mean across items is a simple method for dealing with missing data among scale items. Before adding variables it is a good idea to do a reliability analysis of the scale items.

## Reliability Analysis (Cronbach Alpha)

If you have an additive scale in which you wish to add across multiple items to come up with a score, you must first determine if the items are internally consistent. That is, do the items correlate positively with each other and at a sufficiently high level to justify adding them together to measure the concept that the scale proposes to measure. If you developed your own scale or modified an existing scale this is particularly important, but even if you're using a standardized measure for which the Cronbach alpha (or other measure of internal consistency) was reported, this is still an interesting and instructive thing to do.

1. Select Statistics > Scale > Reliability Analysis. The Reliability Analysis dialog box opens.



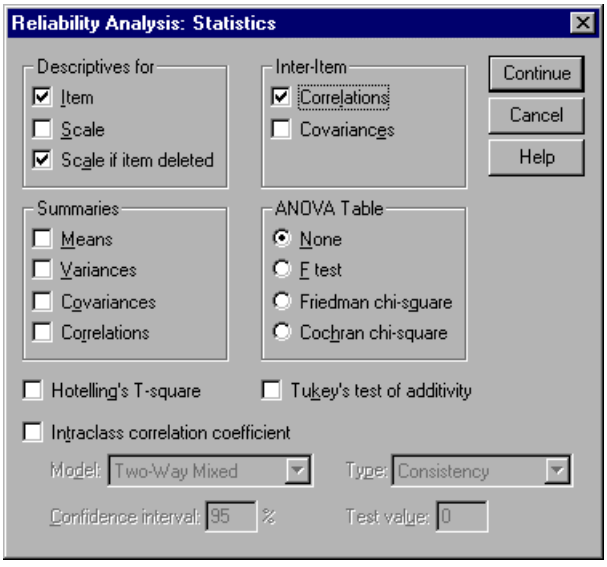

2. Select the "worst state self-esteem" variable names that comprise the scale (**ability1** through **foolish1**) and click on the arrow button so that they appear in the **Items** area.

   Select **Alpha** in the **Model** field.

   Check the **List item labels** option.

3. Click on the **Statistics** button. The Reliability Analysis: Statistics dialog box opens.

4. Select **Item** and **Scale if item deleted** in the **Descriptives for** area, and **Correlations** in the **Inter-Item** area.



5. Click the **Continue** button. The Reliability Analysis: Statistics dialog box closes.
6. Click **OK**. The Reliability Analysis dialog box closes and SPSS activates the output window.

The output gives you the Cronbach alpha statistic. A rule of thumb is the 0.60 is the lower level of acceptability for the alpha, and alphas in the 0.70's and 0.80's are preferable. The inter-item correlation matrix will allow you to see if any of the items are negatively correlated with the other items in the scale, and the "Scale if item is deleted" will reveal the alpha if each item were deleted from the scale. This will help you identify any 'clunkers' in the scale that might need to be removed before the scale score is calculated.

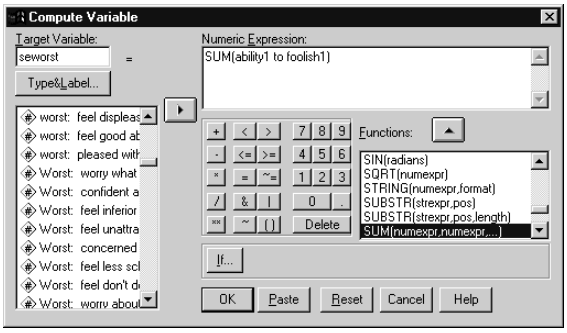
Sum

To sum across variables:

1. Select Transform > Compute. The Compute Variable dialog box opens.
2. In the Target Variable field, enter the variable name **seworst** (for "self-esteem worst"). This variable will be created and the sum entered for each case.
3. In the Functions box select the function **SUM(numexpr,numexpr, . . .)**. Click the up arrow button to add the function to the Numeric Expression box.
4. We want to include each of the self-esteem related items: the variables ability1, success1, up through foolish1. This can be done in one of two ways. You may type in the name of each individual variable, separated by commas. To do this, delete the two question marks that appear between the parentheses and replace them with the individual variable names separated by commas. If we were interested in summing a subscale such as "body self-esteem", we could use the expression **SUM(body1, weight1, appear1, unatt1)**. When you are summing a large number of variables, entering the individual variable names can be tedious and error-prone.

   If the variables appear one after the other in the spreadsheet, as they do in this example, you can simply enter a range of variables to sum. Delete the two question marks that appear between the parentheses and replace them with the phrase **ability1 to foolish1**, so the expression reads:

   **SUM(ability1 to foolish1)**




5. Click OK. The Compute Variable dialog box closes and SPSS activates the data editor. The new variable **seworst** appears in the last column of the spreadsheet.

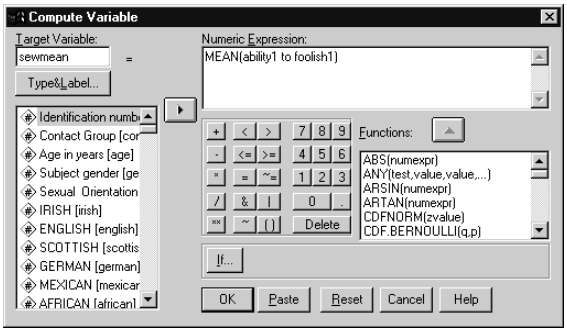
Mean

To compute the mean you use the same dialog box, substituting the MEAN function for the SUM function.

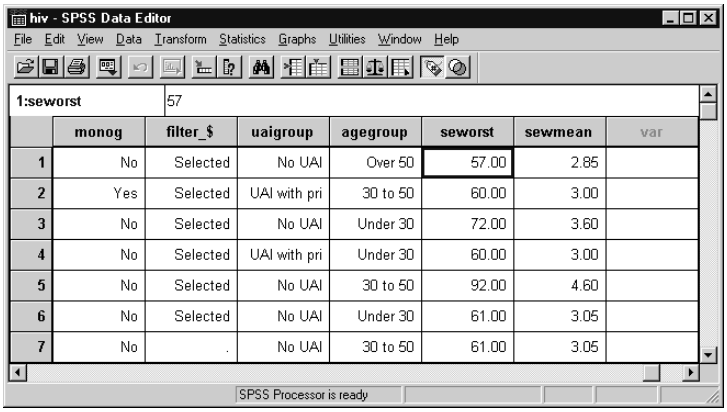
1. Select Transform > Compute. The Compute Variable dialog box opens. If you have previously computed a variable, the data used for that computation is retained in the Compute Variable dialog box. To clear the dialog box in preparation for a new computation, press the Reset button.
2. In the Target Variable box, enter the variable name **seworst** (for "self-esteem worst: mean").This variable will be created and the mean entered for each case.
3. In the Functions box select the function **MEAN(numexpr, numexpr, . . .)** Click the up arrow button to add the function to the Numeric Expression box.
4. We want to include each of the self-esteem related items: the variables ability1, success1, up through foolish1. This can be done in one of two ways. You may type in the name of each individual variable, separated by commas. To do this, delete the two question marks that appear between the parentheses and replace them with the individual variable names separated by commas. If we were only interested in the body image related items, we could use the expression **MEAN(body1, weight1, appear1, unatt1)**. When you are summing a large number of variables, entering the individual variable names can be tedious and error-prone.

   If the variables appear one after the other in the spreadsheet, as they do in this example, you can simply enter a range of variables. Delete the two question marks that appear between the parentheses and replace them with the phrase **ability1 to foolish1**, so the expression reads:

   **MEAN(ability1 to foolish1)**




5. Click OK. The Compute Variable dialog box closes and SPSS activates the data editor. The new variable sewmean appears in the last column of the spreadsheet.




| | monog | filter_$ | uaigroup | agegroup | seworst | sewmean | var |
|---|---|---|---|---|---|---|---|
| 1 | No | Selected | No UAI | Over 50 | 57.00 | 2.85 | |
| 2 | Yes | Selected | UAI with pri | 30 to 50 | 60.00 | 3.00 | |
| 3 | No | Selected | No UAI | Under 30 | 72.00 | 3.60 | |
| 4 | No | Selected | UAI with pri | Under 30 | 60.00 | 3.00 | |
| 5 | No | Selected | No UAI | 30 to 50 | 92.00 | 4.60 | |
| 6 | No | Selected | No UAI | Under 30 | 61.00 | 3.05 | |
| 7 | No | . | No UAI | 30 to 50 | 61.00 | 3.05 | |

## Analyzing the Data Set

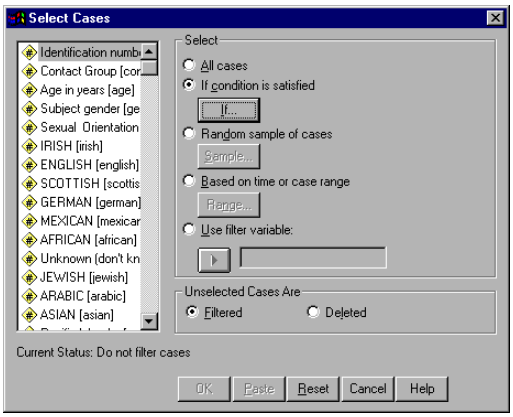
### Analyzing a Subset of Data

You may sometimes wish to exclude some cases while you are performing certain analysis. For example, if you have only a few men in a large set of data, you might decide to select only the females from your population to keep the effect of those few men out of your statistics. In the data set we are working with, the population is predominantly homosexual, with relatively few bisexual subjects. To look just at the homosexual population, without permanently deleting the bisexual cases, we could select just the homosexual cases.

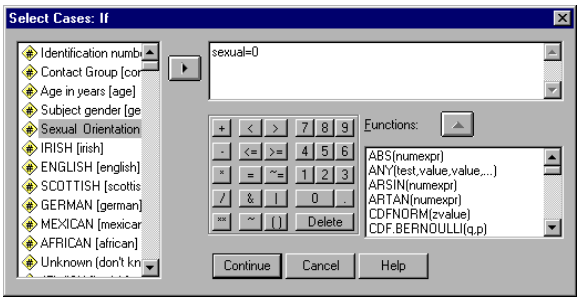
To work with a subset of the data:

1. Select Select Cases from the Data menu. The Select Cases dialog box opens.
2. Make sure the **Filtered** option is selected in the **Unselected Cases Are** area. If **Deleted** is selected the cases will be permanently removed.

   Select **If condition is satisfied** and click on the **If** button. The Select Cases: If dialog box opens.

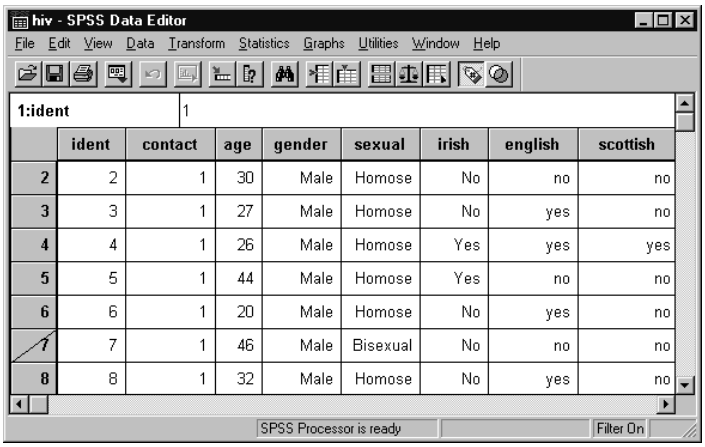



3. Select **Sexual Orientation [sexual]** and click on the arrow button so the variable **sexual** appears in the expression area. Complete the expression by typing in "=0", as 0 is the value for homosexual orientation.




4. Click Continue. The Select Cases: If dialog box closes and you are returned to the Select Cases dialog box.

5. Click OK. You are returned to the SPSS Data Editor. Notice that the bisexual cases still appear in the Data Editor, but there is a slash through the case number, and there is a message on the bottom of the window that there is a **Filter On**.




| | ident | contact | age | gender | sexual | irish | english | scottish |
|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 1 | 30 | Male | Homose | No | no | no |
| 3 | 3 | 1 | 27 | Male | Homose | No | yes | no |
| 4 | 4 | 1 | 26 | Male | Homose | Yes | yes | yes |
| 5 | 5 | 1 | 44 | Male | Homose | Yes | no | no |
| 6 | 6 | 1 | 20 | Male | Homose | No | yes | no |
| 7 | 7 | 1 | 46 | Male | Bisexual | No | no | no |
| 8 | 8 | 1 | 32 | Male | Homose | No | yes | no |

Only those cases that don't have a slash through them will be included in subsequent analyses.

To look only at a consecutive group of cases, or instances if all your control cases are entered consecutively, you can use the Based on Time or

Case Range option in the Select Cases dialog box. See the SPSS Reference Guide for more information on selection criteria.

6. To return to analyzing the complete data set, select Select Cases from the Data menu. Click on the All Cases option, and click OK. Notice that the hash marks are removed from the bisexual cases.

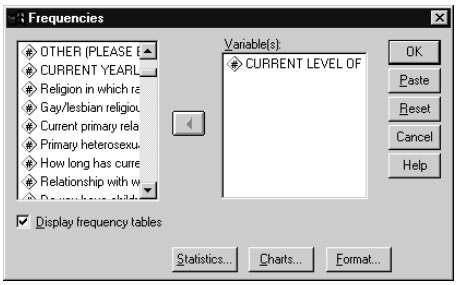
## Descriptive Statistics

Descriptive Statistics look at a single variable. Two commonly used descriptive statistics are frequency distribution, and mean and standard deviation. Frequency distributions are used primarily with categorical data. Mean and standard deviation are used primarily with continuous data.

### Frequency Distribution

This example generates a frequency distribution showing the relative frequencies of education level categories in the hiv data set.

1. Select Statistics > Summarize > Frequencies. The Frequencies dialog box opens.

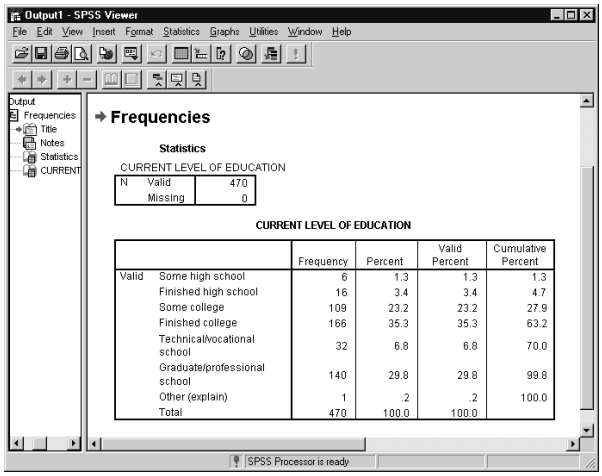



2. Click on the variable **CURRENT LEVEL OF EDUCATION [educate]** in the variable list, then click on the right arrow button between the boxes. The educate variable moves over to the Variable(s) box.

> NOTE: You can calculate multiple frequency distributions simultaneously by adding variables to the Variable(s) box. A single variable is used in this example for clarity.

3. Click on the checkbox next to "Display frequency tables," so a checkmark appears in the box.
4. Click OK.

SPSS activates the output viewer and produces a frequency table for the variable educate.




## Examining and Using the Output in the SPSS Viewer

To examine your output, make the Viewer the active window. You can do this in any of three ways. Select SPSS Viewer from the Window menu; click the SPSS Viewer tab on your Windows taskbar along the bottom of the screen; or if some portion of the Viewer window is visible, click directly on it to make it active.

In the left panel of the Viewer is an outline of the various outputs, listing the name of the analyses and the individual charts or tables produced for each analysis. To go directly to a particular analysis, click on the name in the outline panel. Alternately, you can use the scroll bar on the right panel to scroll up and down through the output.

When the output from a procedure is very long, the Viewer only displays the top portion of the table, then shows a red arrow indicating that the output continues out of view. To view the complete table, double-click on the table itself. A Pivot Table window opens, displaying the complete output. Use the scroll bar to view any part of the table. To close the Pivot Table and return to the Viewer, click on the 'x' in the upper right corner of the Pivot Table window.

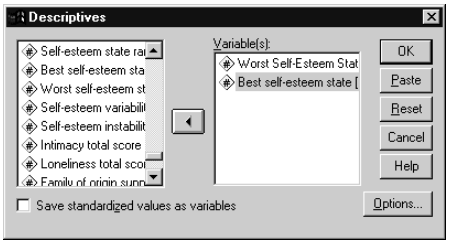
An easy way to create reports from SPSS output is to copy and paste the tables of interest into a word processor. Open both your word processor and the SPSS Viewer. Select a table in the Viewer by clicking directly on it, or by clicking on its name in the outline panel on the left of the Viewer. Once a table is selected, select Copy from the Edit menu in SPSS Viewer, then select Paste from the Edit menu in your word processor.

## Mean and Standard Deviation

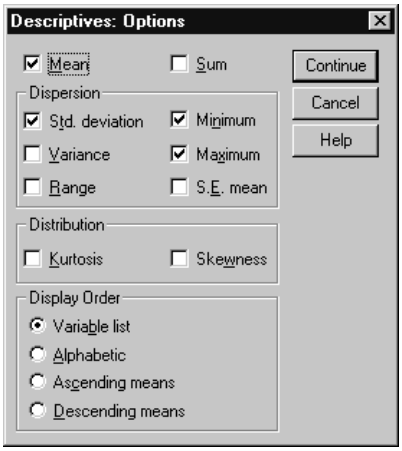
To generate means and standard deviations:

1. Select Statistics > Summarize > Descriptives. The Descriptives dialog box opens.
2. Select **Worst Self-Esteem State [worst]** and click on the arrow button to add the variable worst to the Variable(s) box. Likewise, select **Best self-esteem state [best]** and click on the arrow button.

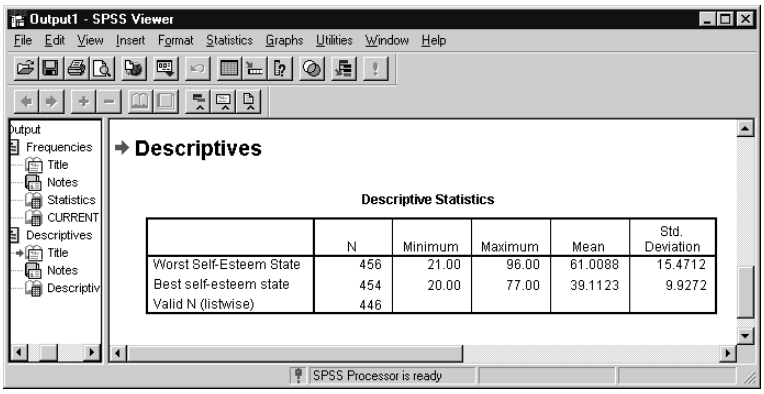
> NOTE: As with frequency distribution, SPSS allows you to calculate mean and standard deviation for a single variable or for many variables simultaneously.




3. Click on the Options button. The Descriptives: Options dialog box opens.
4. Check the Mean option. Select the Std. Deviation, Minimum and Maximum options in the Dispersion area. Select Variable List in the Display Order area, to display the output in the order in which the variables are entered in the Descriptives dialog box.




5. Click the Continue button. The Descriptives: Option dialog box closes and you are returned to the Descriptives dialog box.
6. Click OK. The Descriptives dialog box closes and SPSS activates the output viewer.



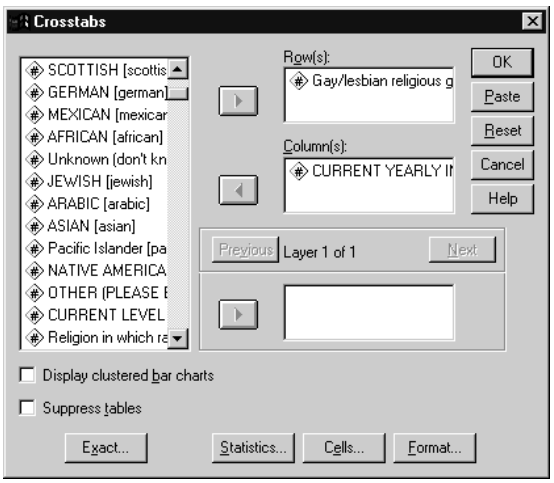

## Comparative Statistics

Comparative statistics look at the relationship between two or more variables. Four of the more commonly used comparative statistics are crosstabulation tables, correlations, t-tests, and analyses of variance.

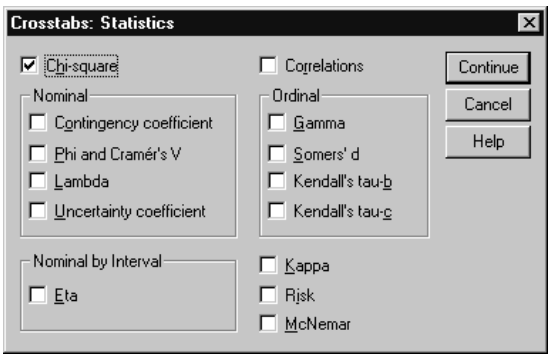
### Crosstabulation Table

To create a crosstabulation table with a chi-square:

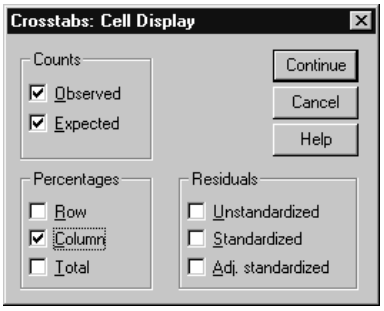
1. Select Statistics > Summarize > Crosstabs. The Crosstabs dialog box opens.
2. Select **Gay/Lesbian Religious Group [galesrel]** and click on the top arrow button to add the variable galesrel to the Row(s) box. This is our dependent variable.
3. Select **CURRENT YEARLY INCOME [income]** and click on the middle arrow button to add the variable income to the Column(s) box. This is our independent variable.



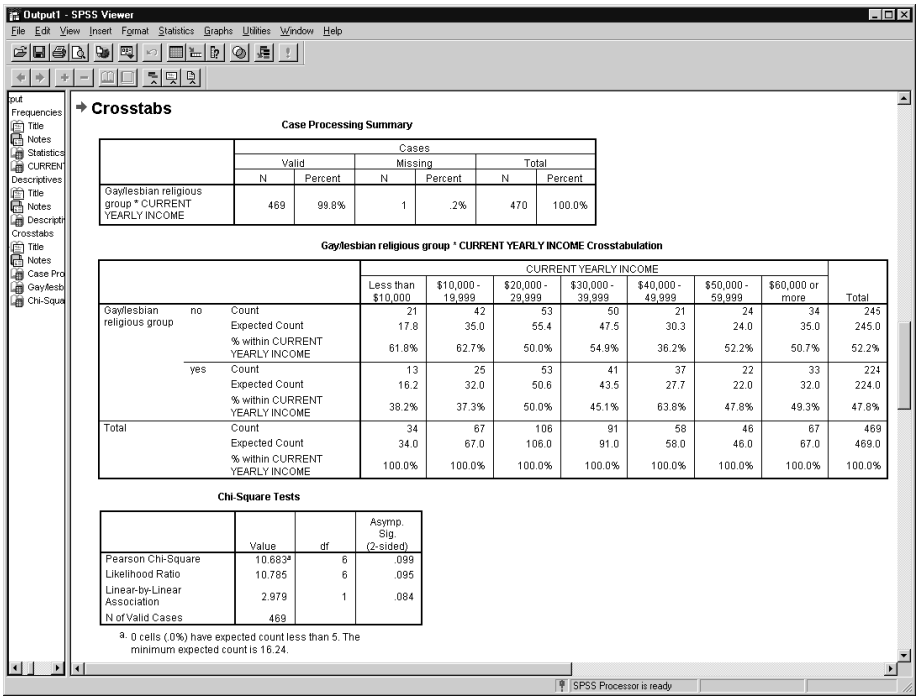

4. Click on the Statistics button at the bottom of the dialog box. The Crosstabs: Statistics dialog box opens.
5. Select the Chi-Square option, so a checkmark appears in the box next to it.




6. Click Continue. The Crosstabs: Statistics dialog box closes and you are returned to the Crosstabs dialog box.
7. Click on the Cells button. The Crosstabs: Cell Display dialog box opens.



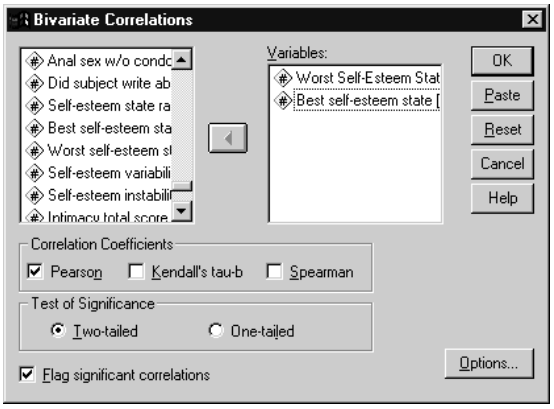

8. Select the Expected and Observed options in the Counts area and the Column option in the Percentages area. The rule of thumb is to produce percentages for the independent variable. In our example the independent variable, income, appears in the column, so we will generate column percentages.
9. Click on the Continue button. The Crosstabs: Cell Display dialog box closes and you are returned to the Crosstabs dialog box.
10. Click OK. The Crosstabs dialog box closes and SPSS activates the output window. The crosstabulation table is displayed first, followed by the chi-square information. Note that SPSS includes a message regarding how many of the cells have an expected count of less than 5.




Correlation

Two or more variables may be included in a correlation matrix. To produce a correlation matrix:

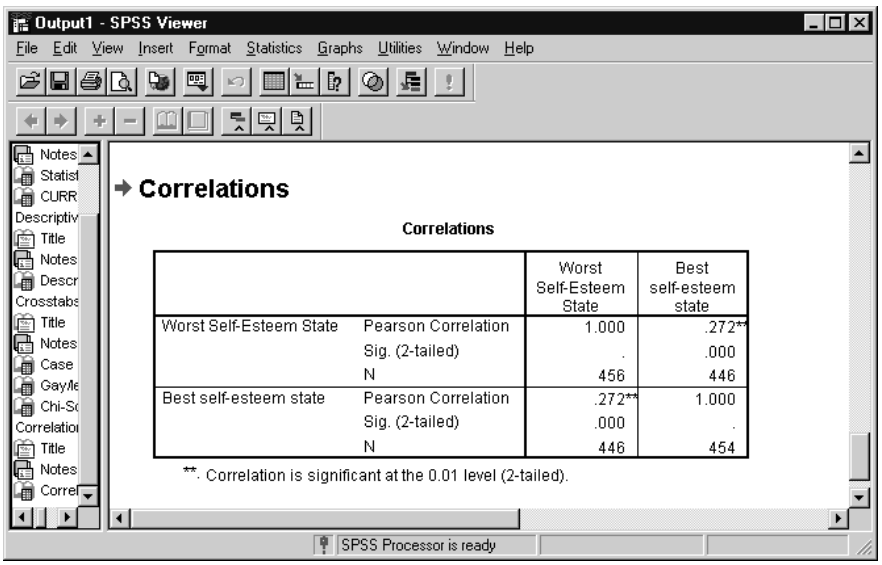
1. Select Statistics > Correlate > Bivariate. The Bivariate Correlations dialog box opens.




2. Select **Worst Self-Esteem State [worst]** and click on the arrow so the worst variable is added to the Variables dialog box. Use the same method to add the variable **Best self-esteem state [best]**. The order of the variables in the Variables box determines the order in which they appear in the report.

3. In the Correlation Coefficients area, select Pearson. In the Test of Significance area, select Two-tailed. Select the Flag significant correlations option at the bottom of the dialog box.

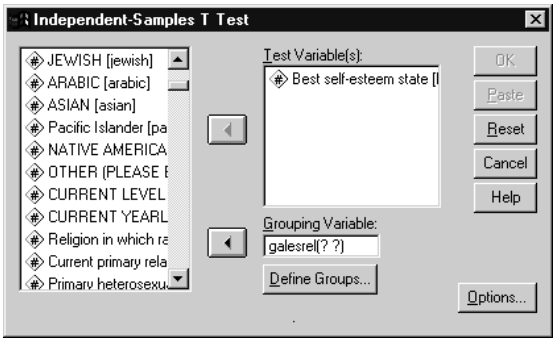
4. Click OK. The Bivariate Correlations dialog box closes and SPSS activates the output viewer. The correlation coefficient for each pair of variables displays in the Pearson correlation area, followed by the two-tailed significance value in the Sig area. The number of cases appears at the bottom.




## Comparison of Means: t-test

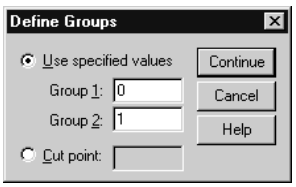
The t-test is a comparison of the means of a dependent variable for two groups of an independent variable. The independent variable must be a categorical variable, or a continuous variable recoded into categorical form. This is called the grouping variable. The dependent variable must be a continuous variable. In SPSS the dependent variable is called the test variable. We will use a t-test to look at the relationship between membership in a gay/lesbian religious group and best self-esteem rating in the hiv data set.

1. Select Statistics > Compare Means > Independent-Samples T Test. The Independent-Samples T Test dialog box opens.
2. Select **Best self-esteem state [best]** and click on the top arrow button, so the variable best is added to the Test Variable(s) box.
3. Select **Gay/lesbian religious group [galesrel]** and click on the bottom arrow button, so the variable galesrel is added to the Grouping Variable box.



4. Highlight the Grouping Variable field and click on the Define Groups button. The Define Groups dialog box opens.

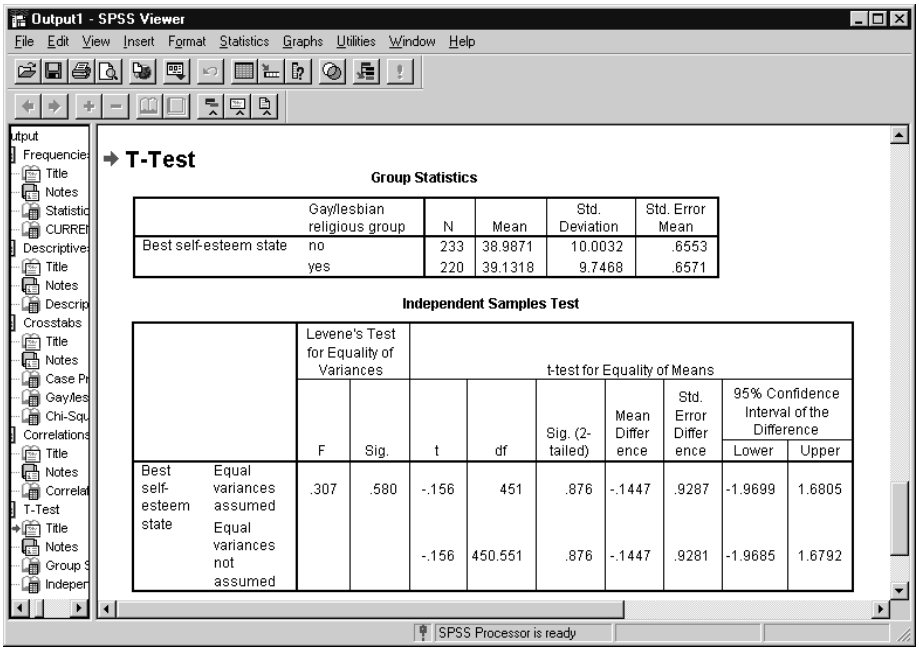
5. Enter **0** (the value for no) in the Group 1 box and **1** (the value for yes) in the Group 2 box.




6. Click on the Continue button. The Define Groups dialog box closes, and you are returned to the Independent-Samples T Test dialog box. Note that the values for the two groups (no and yes) now appear as part of the Grouping Variable expression.

7. Click OK. The Independent-Samples T Test dialog box closes and SPSS activates the output window. A table of means displays, followed by the t-test results.

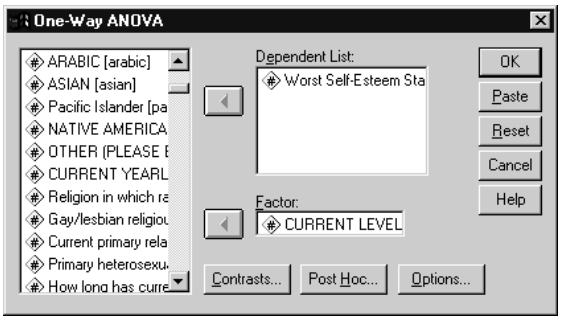
Results of two significance tests are displayed: one for when the variances are equal, and one for when they are unequal. Consult the "p" level for Levene's Test for Equality of Variances in determining which significance test to use. In SPSS, "p" level is known as "Sig" level. If "p" is less than or equal to 0.05, the variances are unequal and the results labeled "Unequal" should be used. If "p" is greater than 0.05, the variances can be considered equal, and the "Equal" results used.



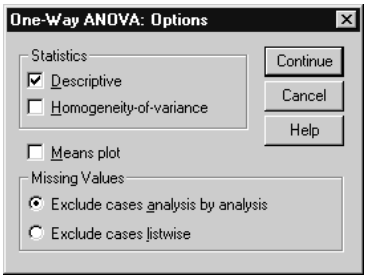
## Comparison of Means: One-Way ANOVA

Like the t-test, one-way analysis of variance compares means of a dependent variable by an independent grouping variable. However, the grouping variable, or "factor," in an analysis of variance may have any number of categories, where the t-test allows only two.

1. Select Statistics > Compare Means > One-Way ANOVA. The One-Way ANOVA dialog box opens. ANOVA is a common abbreviation for Analysis of Variance.
2. Select **Worst Self-Esteem State [worst]** and click on the top arrow button, so the variable worst is added to the Dependent List box.
3. Select **CURRENT LEVEL OF EDUCATION [educate]** and click on the bottom arrow button, so the variable educate is added to the Factor box.



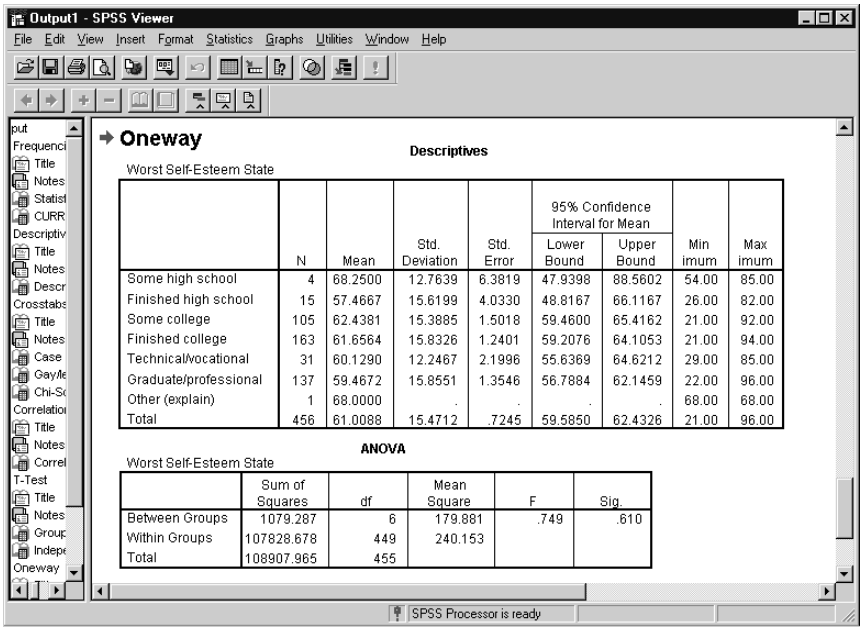

4. Click on the Options button. The One-Way ANOVA: Options dialog box opens.
5. Select Descriptive in the Statistics area, and select Exclude cases analysis by analysis in the Missing Values area.




6. Click on the Continue button. The One-Way ANOVA: Options dialog box closes and you are returned to the One-Way ANOVA dialog box.

> NOTE: Be sure to set the One-Way ANOVA Options (steps 4-6). If you do not, SPSS will use the default options and you will not get the expected report.

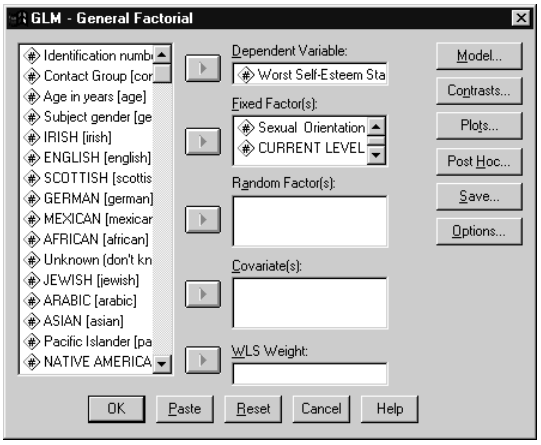
7. Click OK. The One-Way ANOVA dialog box closes and SPSS activates the output window. The analysis of variance significance results are at the bottom right. The means of the dependent variable for each category of the independent variable can be found under “Descriptives”.



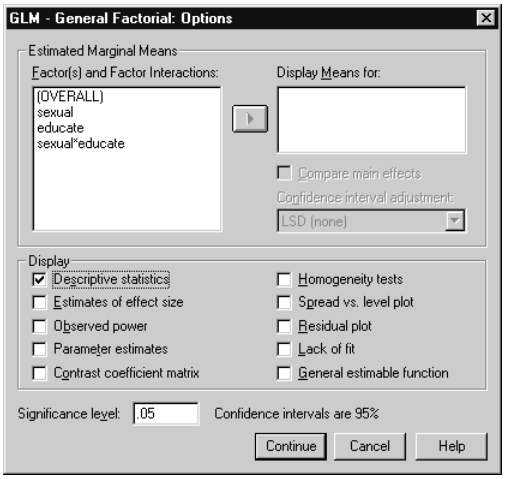

## Comparison of Means: Two-Way ANOVA

The two-way analysis of variance compares means of a dependent variable by two variables or factors.

1. Select Statistics > General Linear Model > GLM - General Factorial. The GLM - General Factorial dialog box opens.



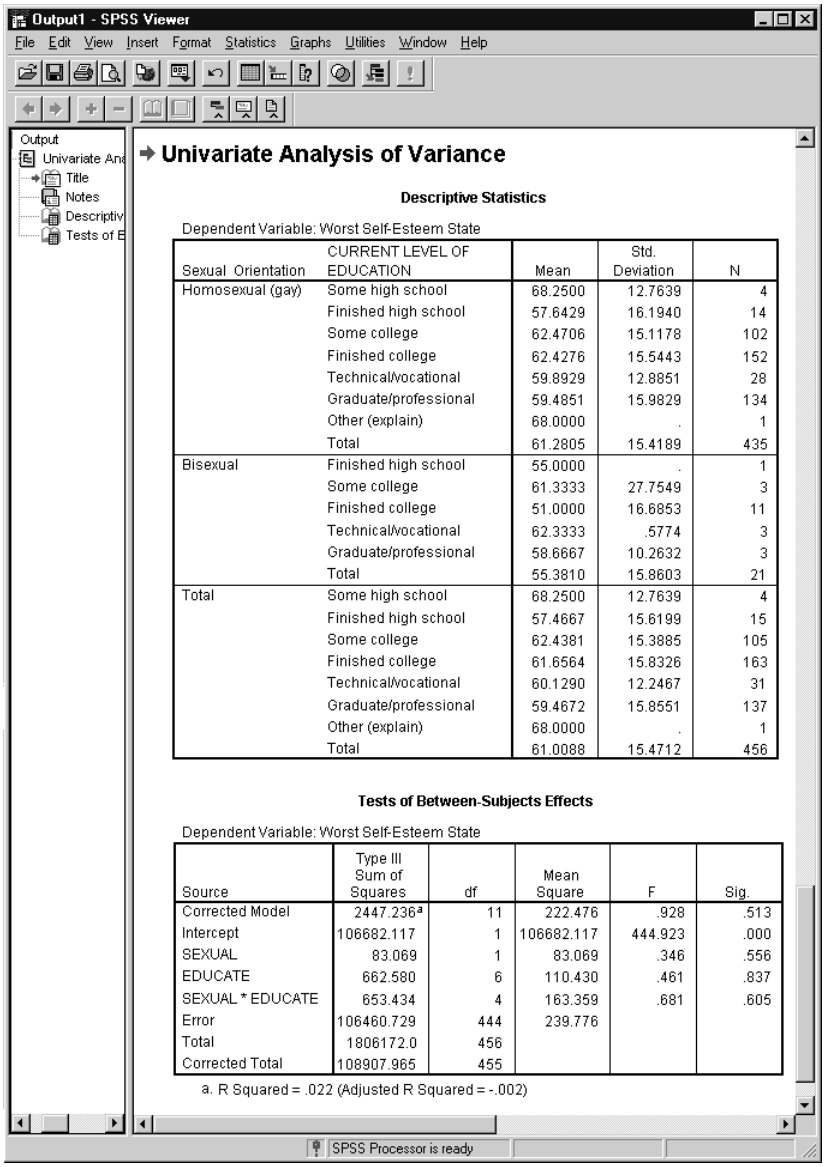

2. Select **Worst Self-Esteem State [worst]** and click on the top arrow button so the variable worst is added to the Dependent Variable box.
3. Select **CURRENT LEVEL OF EDUCATION [educate]** and click on the second arrow button so the educate variable is added to the Fixed Factor(s) box. Next, add the **Sexual Orientation [sexual]** variable to the Fixed Factor(s) box in the same manner.
4. Click the Options dialog box. The GLM - General Factorial: Options dialog box opens.




5. In the Display area, select Descriptive Statistics. Leave the Significance level set to 0.05.
6. Click the Continue button. The GLM - General Factorial: Options dialog box closes, and you are returned to GLM - General Factorial dialog box.

> NOTE: Be sure to set the GLM - General Factorial Options (steps 4-6). If you do not, SPSS will use the default options and you will not get the expected report.

7. Click OK. The GLM - General Factorial dialog box closes and SPSS activates the output window.


Output1 - SPSS Viewer
File Edit View Insert Format Statistics Graphs Utilities Window Help
Output
Univariate Ana
Title
Notes
Descriptiv
Tests of E
Univariate Analysis of Variance
Descriptive Statistics
Dependent Variable: Worst Self-Esteem State
Sexual Orientation | CURRENT LEVEL OF EDUCATION | Mean | Std. Deviation | N
Homosexual (gay) | Some high school | 68.2500 | 12.7639 | 4
Finished high school | 57.6429 | 16.1940 | 14
Some college | 62.4706 | 15.1178 | 102
Finished college | 62.4276 | 15.5443 | 152
Technical/vocational | 59.8929 | 12.8851 | 28
Graduate/professional | 59.4851 | 15.9829 | 134
Other (explain) | 68.0000 | . | 1
Total | 61.2805 | 15.4189 | 435
Bisexual | Finished high school | 55.0000 | . | 1
Some college | 61.3333 | 27.7549 | 3
Finished college | 51.0000 | 16.6853 | 11
Technical/vocational | 62.3333 | .5774 | 3
Graduate/professional | 58.6667 | 10.2632 | 3
Total | 55.3810 | 15.8603 | 21
Total | Some high school | 68.2500 | 12.7639 | 4
Finished high school | 57.4667 | 15.6199 | 15
Some college | 62.4381 | 15.3885 | 105
Finished college | 61.6564 | 15.8326 | 163
Technical/vocational | 60.1290 | 12.2467 | 31
Graduate/professional | 59.4672 | 15.8551 | 137
Other (explain) | 68.0000 | . | 1
Total | 61.0088 | 15.4712 | 456
Tests of Between-Subjects Effects
Dependent Variable: Worst Self-Esteem State
Source | Type III Sum of Squares | df | Mean Square | F | Sig.
Corrected Model | 2447.236a | 11 | 222.476 | .928 | .513
Intercept | 106682.117 | 1 | 106682.117 | 444.923 | .000
SEXUAL | 83.069 | 1 | 83.069 | .346 | .556
EDUCATE | 662.580 | 6 | 110.430 | .461 | .837
SEXUAL * EDUCATE | 653.434 | 4 | 163.359 | .681 | .605
Error | 106460.729 | 444 | 239.776
Total | 1806172.0 | 456
Corrected Total | 108907.965 | 455
a. R Squared = .022 (Adjusted R Squared = -.002)
SPSS Processor is ready

# 4. Additional Sources of Information

**Online Help:** SPSS for Windows includes an extensive online help system. To access the online help, select Topics from the Help menu.

**Trade Books:** SPSS documentation and a variety of third-party trade books are available at the NYU Computer Store.

**Classes:** The Academic Computing Facility offers classes in SPSS and Windows basics. Schedules are available at any ACF computer lab; at the ACF HelpCenter on the second floor of Warren Weaver Hall; and via the World Wide Web at **http://www.nyu.edu/acf/classes/**.

**Consultants:** Researchers can contact Frank LoPresti of the ACF Social Science Statistics and GIS Group at 998-3398 for assistance with the high-level functions of SPSS and other statistical programs.