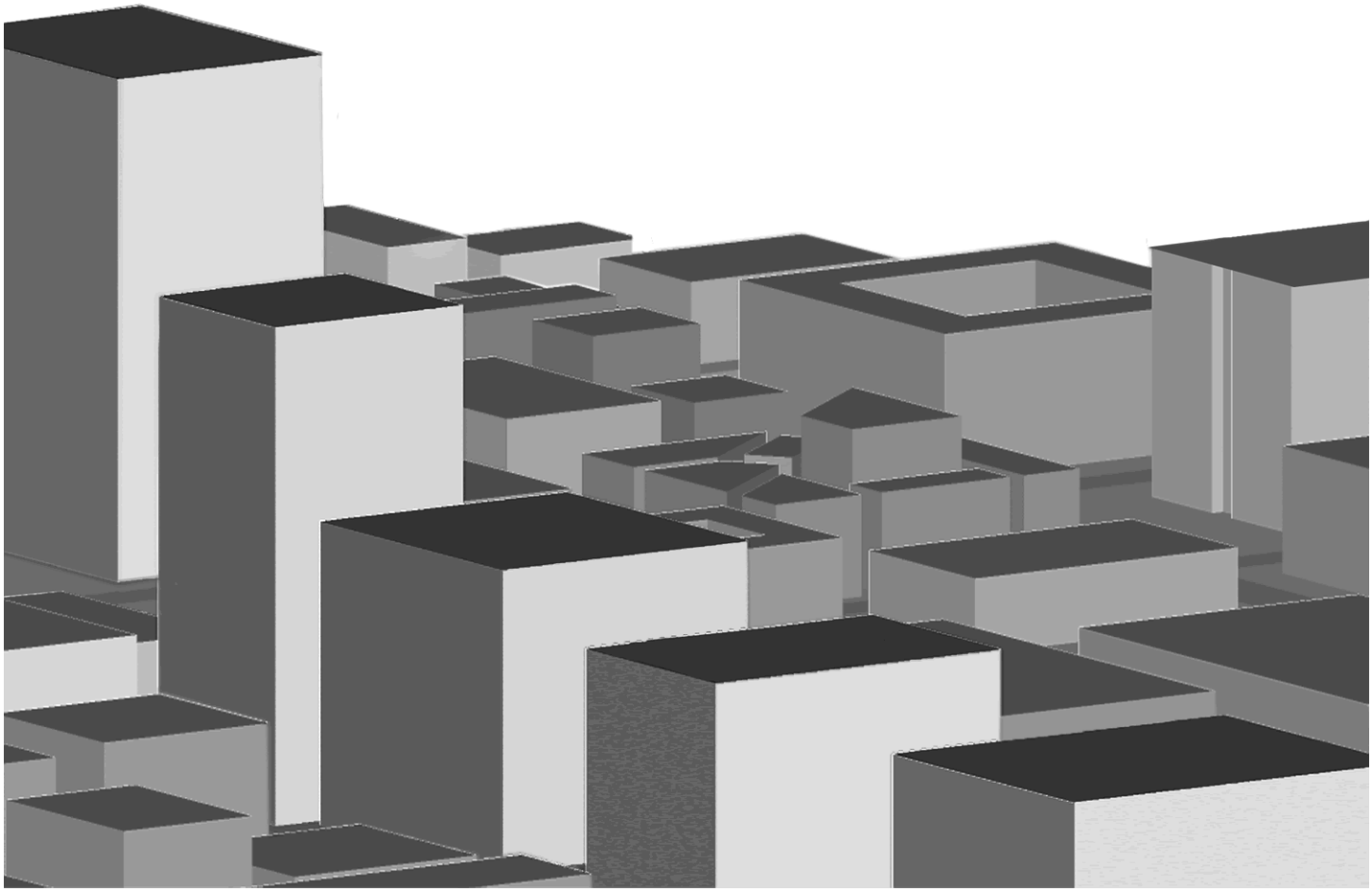
INDIANA UNIVERSITY
UNIVERSITY INFORMATION TECHNOLOGY SERVICES

IT TRAINING & EDUCATION
STEPS TRAINING MATERIALS

http://ittraining.iu.edu
iub: 812.855.7383
iupui: 317.274.7383

SPSS: The Basics



Windows & MacOS

SPSS: The Basics v8.0.0

58 Pages

About IT Training & Education

The University Information Technology Services (UITS) IT Training & Education program at Indiana University offers instructor-led computing workshops and self-study training resources to the Indiana University community and beyond. We deliver training to more than 30,000 participants annually across all Indiana University campuses. Our staff is comprised of enthusiastic professionals who enjoy developing and teaching computing workshops. We appreciate your feedback and use it to improve our workshops and expand our offerings. We have received several national awards for our materials and they are being used at universities across the country. Please keep your questions, comments and suggestions coming!

In Bloomington, contact us at ittraining@indiana.edu or call us at (812) 855-7383.
In Indianapolis, contact us at ittraining@iupui.edu or call us at (317) 274-7383.

For the most up-to-date information about workshops and schedules, visit us at:

http://ittraining.iu.edu/

Copyright 2007 - The Trustees of Indiana University

Contents

Welcome and Introduction

Welcome to *SPSS: The Basics*.

This workshop is designed to acquaint you with the basic use of SPSS. SPSS is a computer program that provides tools for statistical analysis, data management, and presentation. SPSS can perform a wide variety of statistical techniques, such as the T-test, analysis of variance, factor analysis, and multivariate regression to name just a few.

Today we will learn the specifics of using SPSS to complete two types of statistical tests, T-tests and Regression. No prior experience with SPSS is presumed, though a basic knowledge of statistics is required.

What You Should Already Know

You should have already attended *Windows: Basic Computing Skills* and have some additional skills. Specifically, you should:

- Understand basic spreadsheets
- Have a basic understanding of statistical concepts

What You Will Learn

In the course of this workshop, we will be using SPSS to do the following tasks:

- Explore data
- Conduct a T-Test
- Perform regression analysis
- Generate plots
- Use subgroups of data
- Create cross-tabulation

What You Will Need to Use These Materials

To complete this workshop successfully, you will be provided with:

- The use of SPSS 14.x for Windows or SPSS 13.x for MacOS X
- The exercise file **Class.xls**

Getting Started

These materials presume you will begin work from the desktop, and have any required exercise files located in an epclass folder there. For instructions on obtaining the exercise files, see below.

If you need assistance logging on or starting an application, please consult your instructor.

Finding Help

If you have computer-related questions not answered in these materials, you can look for the answers in the UITS Knowledge Base, located at:

http://kb.iu.edu/

Self-Study Training

Want to learn more on your own?

IT Training Online makes self-study computer-based courses available on a wide range of IT topics. You may also purchase STEPS workshop materials to use in learning on your own. To find out more, go to:

http://ittraining.iu.edu/online/

Getting the Exercise Files

Most of our workshops use exercise files, listed at the bottom of page 1 of the materials. In our computer-equipped classrooms, these files are located in the epclass folder, which should already be on the computer desktop. If you are using our materials in a different location, you may obtain the exercise files from our Web site at:

http://ittraining.iu.edu/workshops/files/

Once you are logged on and have the needed files in an epclass folder on your desktop, you are ready to proceed with the rest of the workshop.

Using SPSS for Statistical Analysis

There are two basic kinds of statistics. ***Descriptive statistics*** summarize important features of a data set, like an average score on a test. ***Inferential statistics*** allow us to test hypotheses about a set of data. That is, they allow us to make inferences about population characteristics based on a sample of the population. An inferential statistic, for example, might allow us to test the hypothesis that children who brush with fluoridated toothpaste have fewer cavities than children who don't (a famous hypothesis tested at IU).

The calculations required for even basic statistics can be very cumbersome and tedious to do manually. Many advanced statistical techniques are practically impossible to do without a computer. A statistical software package has much the same effect on data analysis as a word-processing package has on writing. You will sometimes finish quicker, but the primary effect is that the quality of your analysis is markedly improved. Using statistical software, you can get much more information out of your data, and you should be able to do a better job analyzing it than if you were doing the analysis manually. SPSS is one of the oldest and most widely-used statistical software packages, and is among the better ones available.

Launching SPSS

Let's start the SPSS program.

1. Launch SPSS using the Start menu or a desktop shortcut.

   SPSS starts and you see the opening screen, which resembles a spreadsheet.

   You may see a dialog box that would allow you to open an existing data or create a new one. We will not be using that dialog box, so we can close it now.

2. To close the dialog box, if you see it,

   You see the spreadsheet-like SPSS window.

3. To maximize the SPSS window, if it is not maximized,

Click 

Note for MacOS X Users - To maximize the SPSS window, in the upper left-hand corner, Click  .

The SPSS window expands to take over the whole screen.

Getting Familiar with SPSS

SPSS uses a familiar point-and-click windowed graphical interface. You can do all of your work through menus and dialog boxes. Each type of task happens in its own window--data resides in the ***Data Editor*** window, graphs and results of statistical analysis are generated in the ***Viewer*** window, and so forth.

While the Data Editor and Viewer windows share many menu choices, some commands are only available in their respective windows. For example, data-oriented commands (e.g. for recoding an existing variables) are located in the Data Editor window, but commands used to modify the output are located in the Viewer window.

Finding Help

SPSS has an extensive help system which can quickly find information about most SPSS commands. The help systems in SPSS can be accessed throughout the program either from the main menu or by clicking the help button in a window or dialog box. The help system is a typical Help interface with text describing the commands, examples of syntax, and keywords.

In the course of this workshop, we will be using an existing data file in a particular format (Excel spreadsheet, see below) to demonstrate the functions of SPSS. We are interested in looking for information on the different types of data files we can use with SPSS.

1. To access the help menu, on the Menu bar,

Click **Help**, Click **Topics**

The Base System Help dialog box appears. Using this dialog box, you may browse through select topics using the “Contents” tab, or you may search through the help index using the “Index” or “Find” tabs.

Note for MacOS X Users - Skip ahead to Step 3.

2. To go to the index,

   Click the Index tab

   Now we can type the word or words we wish to search for in the help file index.

3. To identify our search string, in the text field at the top of the window, type:

   data files

   You will see the list change to reflect the words you type at the top. The results of the search are listed in hierarchical fashion below. The heading "Data Files" has several related topics immediately following it.

4. ***Note for MacOS X Users*** - To identify our search string, in the search field at the top of the window, type: data file types return, then skip ahead to Step 5.

5. In order to learn more about opening files,

   Double Click opening

   Another dialog box called Topics Found appears because there are many topics associated with the heading.

6. To view the information we are interested in,

   Double Click Data File Types

   The information in the help window changes, outlining the different types of files that SPSS can read. We, however, are finished with the help system.

NOTE: You can get additional information by clicking any of the Related Topics at the bottom of the screen.

7. To close the help window, in the upper right-hand corner,

   Click 

   Note for MacOS X Users - To close the help window, in the upper left-hand corner, Click 

Variables and Cases

A data file must be structured in a particular way so that SPSS can use it. SPSS uses data organized in rows and columns. We call the rows ***cases*** and the columns ***variables***. A case contains information for one unit of analysis, such as a person, an animal, a business, or a jet engine. Variables are the information collected for each case, such as age, body weight, profits, or fuel consumption.

Consider a class of three students, with three test scores each.

Last Name	Test 1	Test 2	Test 3
Brown	70	85	73
Johnson	81	88	90
Zimmer	68	70	65

Here, each row represents the test performance of an individual student. Each column has a different variable, and its value in a particular row depends on the individual student. For instance, the variable Last Name contains characters that name students in the class.

SPSS accepts numbers and characters as data. Each column can contain numbers or characters, but not both.

Manipulating Data with SPSS

SPSS can import data from a variety of formats, and now that you know the data format that SPSS expects, you can retrieve and view a pre-existing data file. A convenient option is to import data from a spreadsheet file. Since many people know how to use spreadsheets, entering data into a spreadsheet file and then importing it into SPSS is an easy way to get data into SPSS.

Typically, after you read the data into SPSS, you will then save the file as an SPSS system file. Once saved, this new file will be in an SPSS-specific format. The advantage of creating a system file is that SPSS can retrieve, manipulate, and perform calculations with such a file much more quickly than it can any of the “foreign” file types, such as a file created by Microsoft Excel.

Importing Data Stored in Spreadsheet Form

Spreadsheet files can be stored in a variety of formats. The sample file that we will be using was created with Microsoft Excel. SPSS automatically translates a number of different formats, so your favorite spreadsheet or database program can be used to provide information.

If you plan on doing data entry with a spreadsheet program and then importing the data into SPSS, there are a few important things to know:

- SPSS can only read files created by certain versions of various other programs so you must check the SPSS documentation to find out if SPSS is compatible with the file you wish to read.
- The data must be organized in a rectangular fashion as described above.
- Any missing values can be left blank and SPSS will automatically treat them as such.
- The variable names should be in the first row of each column and be no more than 64 characters long. Please note that you *cannot use spaces* in variable names and the *names must begin with a letter*. Numbers can only be used after the first letter (e.g. "v1" is valid, "1v" is not).

NOTE: Currently, only newer versions of SPSS (SPSS 12 and higher) can use such a long variable name; earlier versions of SPSS have an 8 character limit for variable names. The length of variables names in most other statistics applications is also limited to 8 characters, so if you expect to use the data file elsewhere, or share it with another researcher, current best practice is to use short variable names.

Opening a File

The process for retrieving a "foreign" file for analysis and opening an SPSS system file are nearly identical. The only difference is that the file format must be set correctly for SPSS to be able to properly read the file into memory. In the example below we will read in the exercise file named **Class.xls**.

1. To import the file, on the Menu bar,

You see the Open dialog box.

We need to specify the name and location of the file to open.

Setting the Location for Opening Your File

When the dialog box opens, it lists a default location from where the file will be opened. This location is displayed in the Look in field.

On a Windows PC, the Look in field is located at the top-left corner of the dialog box.

All of our exercise files are contained in the epclass folder, located on the desktop. We'll want to change our Look in location to this folder.

1. To move to the epclass folder, on the Look in drop-down list,

Click , Click epclass

The epclass folder is now listed as the Look In location.

Note for MacOS X Users - On a Macintosh, the Look in field is above the list of visible folders and files, and is labeled "From:". To move to the desktop, press the key combination command + D. From there, you can move into the epclass folder.

By default, SPSS will display only folders and SPSS native files. We must change the file type to Excel spreadsheet in order to see the spreadsheets (specifically **Class.xls**) in the epclass folder.

2. To change the file type to Excel, in the Files of type drop-down menu,

Click the Files of type drop-down list, Click Excel (*.xls)

Note for MacOS X Users - Use the Show: field to allow viewing of Excel files.

You see the list of Excel spreadsheets in the epclass folder.

3. To open the file for this exercise,

Double Click Class.xls

You see the Opening File Options dialog box.

We must now tell SPSS to read the variable names that are located in the first row of each column in the data file.

4. To read in the variable names from the data file,

Click the "Read variable names" checkbox

NOTE: Newer versions of SPSS (SPSS 12 and higher) will now preserve the display of the character case for variable names as they are imported. This means the expected case from the original source will be shown in dialog boxes, the Viewers, and many other places in the SPSS interface, but this case-sensitivity does not extend to the actual syntax of SPSS commands.

5. To finish opening the data file,

You see the SPSS Viewer window open with information regarding the executed command. Whenever a command is executed in SPSS for Windows, the Viewer window is automatically brought to the forefront after command has finished running.

Note for MacOS X Users - While the Windows version of SPSS has duplicate Menu bars above the Data Editor and Viewer windows, the Mac OS X version only has a single Menu bar above the Data Editor.

6. To switch back to the SPSS Data Editor, on the Menu bar,

Note for MacOS X Users - To switch back to the SPSS Data Editor, on the Menu bar, Click Window, Click Untitled - SPSS Data Editor.

What the Sample Data Represent

The data you are looking at are from a study to investigate computer anxiety in middle-school children. The data were collected from forty ninth-graders from three different school systems.

The information collected on each student includes:

- Identification (id) number
- Gender of the student
- Amount of previous computer experience
- School system
- Responses from two ten-item questionnaires that show computer and math anxiety
- Test scores from a computer course and a mathematics course for a given testing period
- Two variables that are measures of computer and math anxiety, based on the questionnaire responses

Each row of data is a single case; that is, each row of data tells us about an individual student. Each column is a single variable. In SPSS, we are required to give each variable a name. Variable names are limited to 64 characters by some versions of SPSS (SPSS 12 and higher), but standard practice is to use variable names consisting of eight characters or fewer.

NOTE: Variable names in SPSS are not case-sensitive in command syntax, but will be displayed using the case in which they were entered or imported. By way of convention, uppercase names are used in these materials to refer to variables.

The first column of data, ID, contains the student id number, a number that the researcher has assigned to each student. The second column, SEX, identifies the student's gender. The third column is the variable EXP, which tells us roughly the number of years of previous computer experience.

- 1 = one year or less
- 2 = about two years
- 3 = three years or more

The next column, SCHOOL, contains numbers that identify which of three school systems the student attends.

- 1 = rural school system
- 2 = suburban school system
- 3 = city school system

The following ten columns, C1 to C10, are the results of a questionnaire designed to test computer anxiety. The next ten columns, M1 to M10, are the results of a questionnaire designed to probe math anxiety.

The next two columns, MATHSCOR and COMPSCOR, are scores from tests given in a math course and a computer course, respectively, for a given time period.

The questionnaire was designed so that responses close to 1 indicate higher anxiety about a given subject, and responses close to 5 indicate lower anxiety. By summing all of the responses for, say, math anxiety, we get a number that gives us a measure of total math anxiety.

MANX, a dichotomous variable, was set to 0 if the sum of the responses were high (greater than or equal to 30 - indicating low math anxiety), and 1 if the student's responses were low (less than 30 - indicating high math anxiety).

CANX is simply the sum of the responses to the computer anxiety questions. A larger value of CANX for a particular student means lower computer anxiety.

Editing the SPSS Data File

The Data Editor window is the place where you can enter new data, edit existing data values, and delete data. It provides two ways to view your data, the Data View and the Variable View. The Variable View allows modifications to variables and attributes of variables, such as data types, number of digits or characters, descriptive labels, and so on. We will see more about this feature shortly. The Data View resembles a spreadsheet, with each row holding the data for a case, the columns are the variables, and the individual cells containing the actual values of a variable for a case.

Let's see how to use the Data View.

Navigating in the Data View

In the Data View of the Data Editor window, you can use the tab, arrow keys, and page up/down keys to scroll to any data value. The environment behaves much like a standard spreadsheet program. You can also use the mouse to click on a variable name or data cell. While viewing this window, it is possible to make changes to the data.

Try moving around in the Worksheet by doing the following:

1. To move the active data cell, press:

   The highlighted region moves right one cell.

2. To see how to move in the worksheet, press:

   Notice the movement of the cursor around the data table.

3. To see how data scrolls, press:

 eight or nine times.

   You see the data scroll in from the right, and out to the left.

4. To see it scroll off the screen, press:

You see the data scroll out of view. In this sense the Data Editor Worksheet reacts like a spreadsheet program, however, the Data Editor Worksheet is not a spreadsheet. You cannot, for example, enter formulas into these cells; you may only enter numbers and letters.

5. To jump to the first variable for the current case, press:

You see the value for that variable is selected.

NOTE: Pressing End will jump to the last variable for the current case.

Editing Data in the Data View

In the Data View of the Data Editor window, you can change data values and variable names by clicking the cell, typing the new value, and pressing Enter. The old value disappears and the new one replaces it.

We'll try this out by assuming that we know that there is a problem with case #17. In this data set, the student was accidentally listed as female when in fact the student is male. We need to correct the data.

1. To change the data, in row 17, under the variable labelled "sex",

You see the value "F" highlighted.

2. To correct the data, type:

Note for MacOS X Users **-** In this workshop, the use of Enter or return are interchangeable.

You see the corrected value entered.

NOTE: While variable names are not case sensitive, SPSS data is case sensitive. SPSS would read "m" as a different value than "M".

Notice that the values for the ID variable are showing two decimal places. It is unlikely that the values for this particular variable need to be displayed with this degree of accuracy, so let's change the number of decimal places that will be displayed. Since this is an attribute of the variable, we will do this using the Variable View feature.

Editing Data in the Variable View

As mentioned previously, the Variable View allows you to add or delete variables, or change attributes of variables.

Changing Displayed Decimal Places

Let's see how to change the number of decimal places for the variable, ID. This change affects only the display of the data, and does not actually change the values in the data set.

First, we'll switch to the Variable View of the Data Editor.

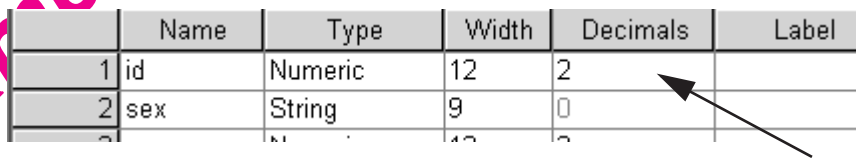
1. To move to the Variable View,

   Click the Variable View tab

   The variables in the data set are now listed in rows, with the columns representing attributes for those variables. Now, let's change the decimals attribute for the ID variable. We will have to activate a hidden spinner control in the decimal attribute cell to do this.

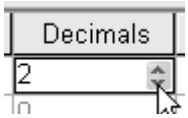
2. To activate the spinner control for the ID variable,

   Click in the Decimals cell



   The spinner control appears on the right-hand side of the cell. We want to change the value to be zero, so no decimal places are displayed.

3. To change the decimals value,

Click 

twice

The decimal value for the ID variable is changed to zero.

Let's also reduce the number of decimal places being displayed for the variable EXP. Since this is a numeric code representing the number of years of previous computer experience we probably do not need to see two decimal places of accuracy for this data. Let's reduce the displayed decimals for this variable to zero, as well.

4. To reduce the displayed decimal places for the EXP variable,

Repeat steps 2 & 3

Adding Value Labels to a Variable

Using numeric codes to represent nominal data is efficient for SPSS to process, but can be taxing for humans to have to remember or decipher the codes when viewing the output of an analysis procedure. SPSS allows the use of labels to be associated with values, which can make the output of some analyses easier to read and understand.

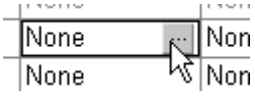
Let's see how to add the value label attributes to the EXP variable. First, we'll have to activate the button which will open the appropriate dialog box.

1. To activate the button in the Values cell for the EXP variable,

Click in the Values cell

The button appears on the right-hand side of the cell. This button will open the Value Labels dialog box.

2. To open the Value Labels dialog box,

Click 

The Value Labels dialog box opens, and the cursor is blinking in the Value field. Let's add the label for the first value, 1.

3. To enter the value, type:

1 Tab

Now we need to provide the label to be used for that value.

4. To enter the label for the value 1, type:

One year or less Click Add

The value/label pair is shown in the list window of the dialog box. SPSS automatically places the label inside of double-quotation marks. Now we'll add the label for the next value, 2.

5. To enter the value, type:

2 Tab

Now we need to provide the label to be used for this value. In this dialog box, pressing the Enter key has the same effect as clicking on the Add button.

6. To enter the label for the value 2, type:

About two years Enter

The second value/label pair is added to the list window of the dialog box. Now we'll add the final label for the last value, 3.

7. To enter the value, type:

3 Tab

Now we need to provide the label to be used for this value.

8. To enter the label for the value 1, type:

Three years or more Enter

The final value/label pair is entered into the list window. Now we can accept the list we entered.

9. To accept the list of value/label pairs,

The dialog box closes, and the beginning of the list is now visible in the values cell for the EXP variable. This process we followed is simple, and sufficient if you already know which values exist, or if you do not have a large number of values. Fortunately, for those other instances, SPSS provides the Defining Variable Properties feature.

Defining Variable Properties

The Defining Variable Properties feature will scan the actual data values, and provide a list of unique values, to make labeling them easier and more accurate. It also allows other variable attributes to be modified, such as changing the decimal places. Let's see how to use this feature to change the displayed decimal places for and add values labels to the SCHOOL variable. First, we'll have to select that variable and all of its attributes.

1. To select the proper variable, on the Menu bar,

The Define Variable Properties dialog box opens, which will allow variables to be selected by name.

NOTE: The controls at the bottom of this dialog box may be useful if you have a very large data set, or one with many variables.

2. To select the SCHOOL variable,

3. To move SCHOOL into the right-hand box,

This action moves the variable from the variable list on the left to the Variables to scan list in the right-hand box.

4. To confirm the selection of the desired variable,

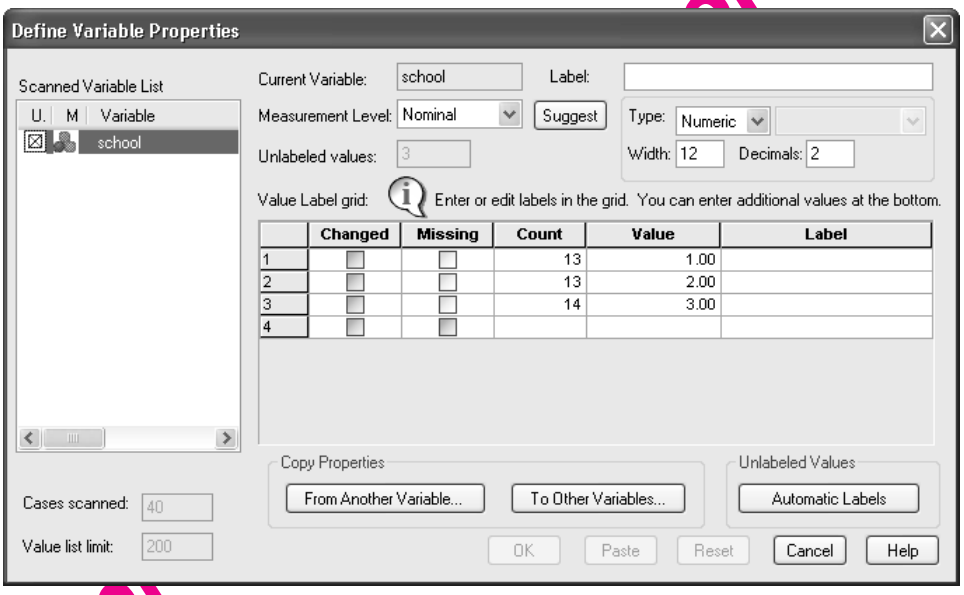
Click Continue

The Define Variable Properties dialog box opens. All fields and controls are currently greyed out, since no variable is selected.

5. To select the SCHOOL variable, in the Scanned Variable List field,

Click school

The dialog box should now look like:



Changing Decimal Display and Value Labels

In the Variable View tab, changing the decimal display and the value labels was tedious and changes were difficult to see at a glance. In this dialog box, the controls are more readily accessed and the results are more apparent. Let's change the displayed decimal places for this variable, first.

1. To select the value for the decimal places,

Double Click the value in the decimals field

The current value, 2, is highlighted.

2. To change the value for the decimal places displayed, type:

0

The value is changed, and all of the buttons at the bottom of the dialog box are now active. Now let's add the appropriate labels for each of the variable's values.

3. To add the label for the first value, 1,

Click in the label field for that value, type: Rural school system

4. To add the label for the second value, 2,

Click in the label field for that value, type: Suburban school system

5. To add the label for the final value, 3,

Click in the label field for that value, type: City school system

The value labels appear next to the values. We'll change one more attribute of this variable, the measurement level.

Changing the Measurement Level of a Variable

The measurement level of a variable describes its data, and may affect how some procedures in SPSS handle the variable. For instance, some procedures may treat categorical variables (Nominal or Ordinal) differently than scale variables (Interval or Ratio).

NOTE: In ascending order of sophistication, the traditional taxonomy of measurement levels are: Nominal, Ordinal, Interval and Ratio. In SPSS, the Interval and Ratio levels are grouped together as Scale.

SPSS 14 will assign measurement levels to imported numeric data using the following rule; variables with fewer than 24 unique values and string variables are set to Nominal, while variables with 24 or more unique values are set to Scale. When we imported the data file, the SCHOOL variable was assigned a measurement level of Scale. Remember that this variable is using a numeric code to represent the three types of school systems, so this is really a categorical variable. But should this be considered a Nominal or an Ordinal variable?

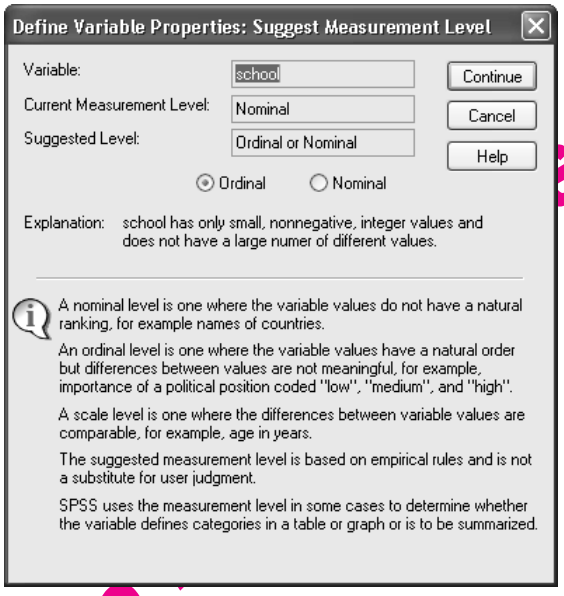
Note for MacOS X Users - SPSS 13.x will assign the Scale measurement level to all imported numeric variables.

SPSS provides a nice feature to help you determine the measurement level in this Define Variable Properties dialog box. SPSS can examine the data in the selected variable, and offer some advice based on traditional empirical rules regarding measurement levels.

1. To receive a suggested measurement level for the selected variable,

Click Suggest

The **Define Variable Properties: Suggest Measurement Levels** dialog box appears:



SPSS offers a suggestion of using Ordinal for SCHOOL. Since there is no ranking intended by the designation of numeric codes for the three categories, we will assign the Nominal measurement level to SCHOOL instead.

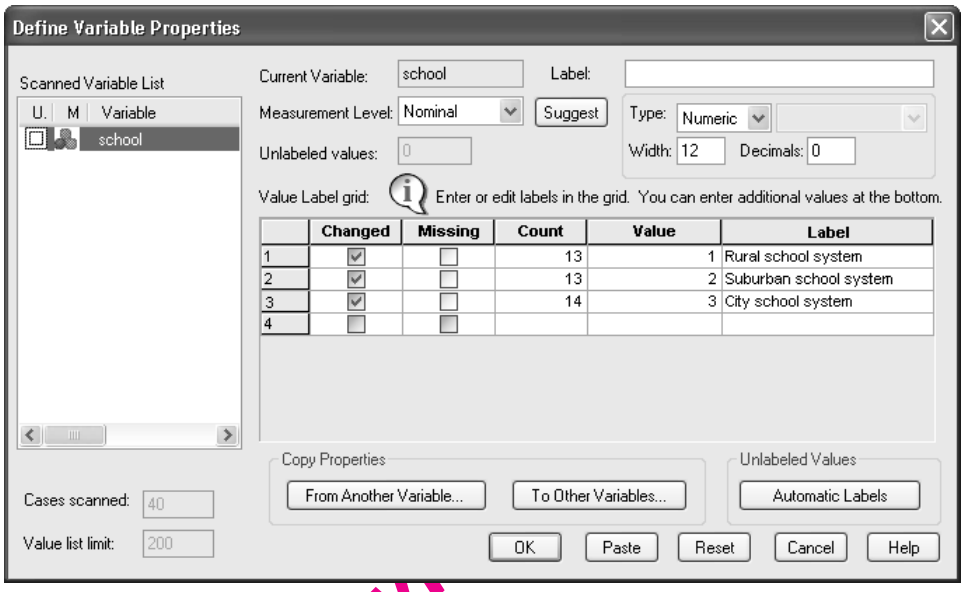
2. To indicate the desired measurement level,

Click the Nominal radio button

3. To accept this change in measurement level,

Click Continue

The completed dialog box should now look like:



4. To apply all of the changes to the SCHOOL variable data,

Click OK

The changes can be seen in the SCHOOL variable row. Let's switch back to the Data View of the Data Editor, to see what these changes look like in that view.

5. To move to the Data View,

Click **the Data View tab**

The changes to the decimal places being displayed are most evident.

We have made a number of changes to the data set, and now should save them. Let's save the edited file as an SPSS system file.

Saving Your Data File

While it is possible to save the file in many different formats, it is best to use the SPSS system format. This is always the default in the Save As dialog box after non-SPSS data is imported and the file is subsequently saved.

Even though we opened this file from disk, since it is an Excel file, SPSS *copied* its data into the new file we are currently working with. This new file has not been saved.

1. To save your data, on the Menu bar,

   Click File, Click Save

   The Save Data As dialog box appears.

2. To name and save the file, in the File Name box, type:

   class Enter

   SPSS will automatically assign the extension ".sav" to the file name so you will know in the future that this is a SPSS data file.

NOTE: When we hit Enter, it performed the same function as clicking the Save button in the dialog box. A highlighted button is always the default option, so you can simply press Enter to continue rather than having to Click with the mouse.

Exploring the Data

One of the advantages of SPSS is that most SPSS data analysis can be accomplished using menus and dialog boxes. Newer versions give us this advantage. In older versions of SPSS, users had to type in all commands manually. We will use the windowed graphical interface features in this workshop.

Descriptive Statistics

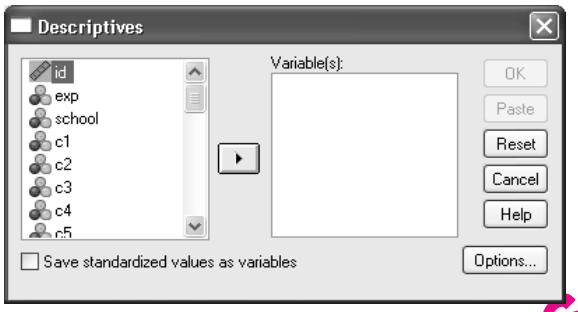
All of the different types of statistical analyses available in SPSS are located under the main menu option "Analyze." In this exercise, we are going to use SPSS to calculate some basic descriptive characteristics of a data set using the "Descriptives" command.

1. To select Descriptive statistics, on the Menu bar,

Click Analyze, Point Descriptive Statistics,

Click Descriptives...

You see the **Descriptives** dialog box:



The Descriptives dialog box is a good example of many of the dialog boxes you see when using SPSS. On the left side of almost every dialog box, there is a text box listing all of the variables in the data set. On the right side, you have one or more empty boxes. To choose variables for statistical analyses, you select variables in the left-hand box and then click the arrow(s) on the right to move them into the right-hand box.

Now we must select variables from the variable list and choose the descriptive statistics we want SPSS to calculate. We want statistics for the variable MATHSCOR.

2. To select the MATHSCOR variable, in the left-hand box,

scroll as needed, Click mathScor

3. To move MATHSCOR into the right-hand box,

Click ▸

This action moves the variable from the variable list on the left to the Variable(s) list in the right-hand box.

NOTE: Other variables can be selected and moved in the same manner. In this exercise, we will choose only the MATHSCOR variable.

4. To generate the default set of descriptive statistics for the MATHSCOR variable,

The results appear in the Viewer window.

Note for MacOS X Users - The Viewer window may be hidden by the Data Editor. To switch to the Viewer window, on the Menu bar, Click Window, Click Output 1 - SPSS Viewer.

The Descriptive Statistics table inside the right-hand pane of the Viewer window presents the results of the statistical calculations. We can see that there are forty cases (N = 40), that the lowest score on the test in the math course was 20 (Minimum = 20.00), and that the highest score was 99 (Maximum = 99.00). The average score on the test was 48.73 (Mean = 48.7250). In the last column, we see the that the standard deviation, a measure of variance, is 16.08 (Std. Deviation = 16.0751).

NOTE: ***Variance*** is a measure of the spread of scores in a distribution of scores; that is, a measure of dispersion. The larger the variance, the further the individual scores are from the mean. ***Standard deviation*** is the square root of the variance. In a normal distribution, about 99% of all scores should fall within three standard deviations of the mean score. Here, three standard deviations either way from the mean gives a range of scores 48.73 ± 3 x (16.08); the scores from 0 to 96.97.

Often, just looking at a table of descriptive statistics doesn't help us understand the data all that much. In many cases we will want to plot the data to get a visual image of the nature of the data as well.

Plotting Data

Generating a ***scatterplot*** is often used as an intermediate step to check the data for anomalies or data entry mistakes. Because a scatterplot graph will allow you to visually inspect all of the data at once, this is a quick way of checking for obvious trends, outliers, etc.

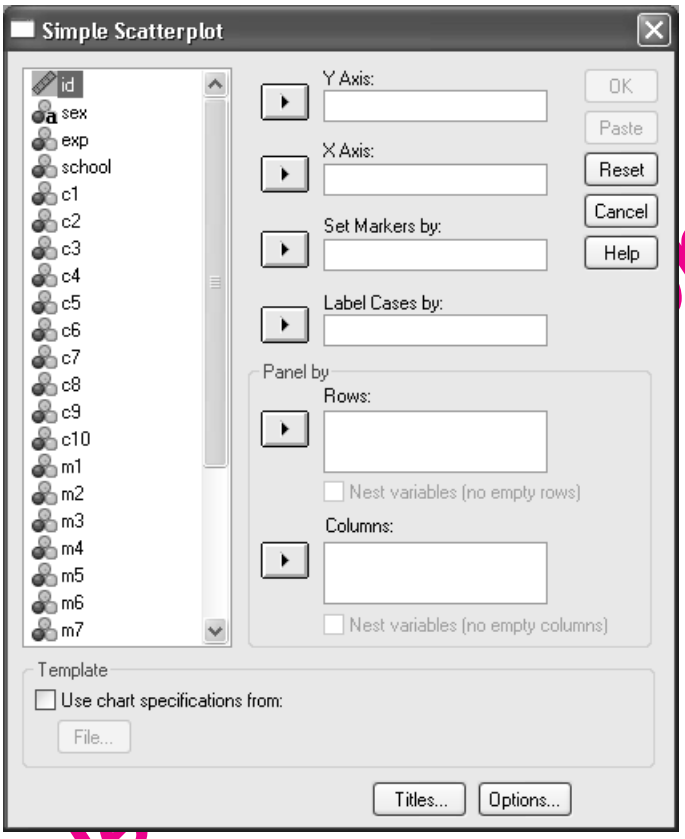
1. To obtain a scatterplot graph of the variable MATHSCOR by each student, on the Menu bar,

The Scatterplot dialog box appears. We are interested in the simple scatterplot, which is the highlighted default option.

2. To choose the default simple scatterplot,

Click Define

You see the **Simple Scatterplot** dialog box:



3. To select the MATHSCOR variable, in the variable list,

scroll as needed, Click mathScor

4. To identify MATHSCOR as the Y axis (vertical axis), next to the box labelled "Y Axis",

Click ▸

The MATHSCOR variable moves to the Y Axis box.

5. To select the ID variable for use as the X axis variable,

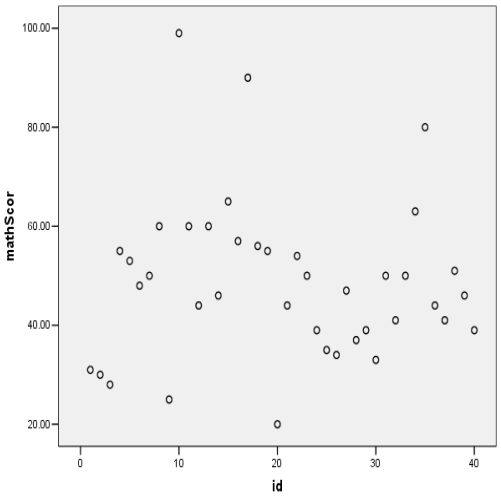
scroll as needed, Click id

6. To identify ID as the X axis (horizontal axis), next to the box labelled "X Axis",

The ID variable moves to the X Axis box.

7. To generate the scatterplot,

You see the plot in the Viewer window, showing the test score dispersion.



As you can see, the distribution is relatively even although the minimum and maximum scores are well above and below the mean.This would be more evident if there were a marker for the mean of the MATHSCOR in the plot.

To create a mean line for the MATHSCOR data points, we must edit the scatterplot.

8. To edit an existing graph,

Double Click **on the scatterplot**

The Chart Editor window opens. This is the edit mode of the SPSS Chart engine where pre-existing graphs can be changed (e.g. text added, lines added, markers changed).

Now we want to make SPSS fit a mean line to the data points. First, we'll select all the data points.

9. To select all of the data points in the plot,

Click *any data point*

The data points are selected. Now we can have SPSS fit a line at the mean point of the selected data points.

10. To fit a line to the selection, on the Menu bar,

Click **Elements,** Click **Fit Line at Total**

The Properties dialog box opens. We want to have a line inserted at the mean point on the Y axis.

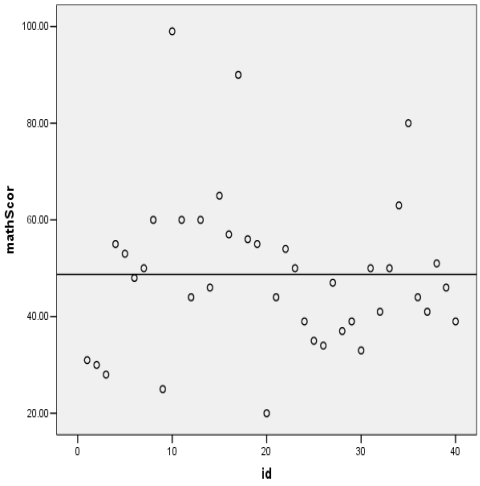
11. Make certain the Fit Line tab is active.

12. To insert the mean line,

Click the "Mean of Y" radio button, Click Apply, Click Close

Note for MacOS X Users - To insert the mean line, Click Chart, Click Options..., and the Scatterplot Options dialog box opens. Then Click the Total checkbox in the section titled "Mean of Y Reference Line", and Click OK.

You see the mean reference line appear on the scatterplot with the raw data.



13. To close the Chart Editor and return to the Viewer window,

Click 

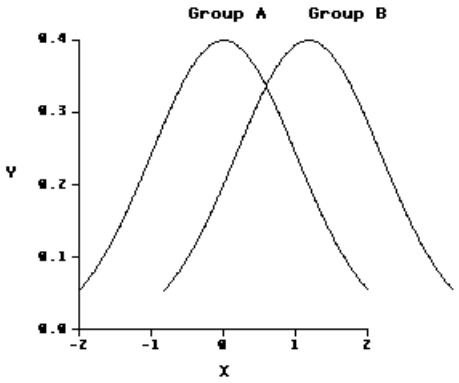
Note for MacOS X Users - To close the Chart Editor and return to the Viewer window, Click  .

Now the relationships between the scores and the mean is more apparent.

Performing a T-test

A ***t-test*** is an inferential statistical analysis used to decide whether two population means are the same or different. Although the theory of t-tests is rather involved, a brief explanation will help us understand how we can use SPSS to perform a t-test.

Suppose that we have two samples labelled “Group A” and “Group B” with distributions shown as follows:



By looking at the mean and the variance of each sample, we can make an inference about the likelihood that these two samples are from two populations with the same mean. The ***null hypothesis*** is that the two samples come from populations with the same mean. The ***alternative hypothesis*** is usually that the two samples come from populations with different means.

With a t-test, we calculate the probability of obtaining two sample values as far apart or farther apart than the observed values if the null hypothesis is true. The typical interpretation is that if the probability of getting two sample means at least as far apart as those observed is 5% or less, then we conclude that the results are so unlikely under the null hypothesis that the null hypothesis is not true. That is, we reject the null hypothesis that the samples were drawn from populations with the same mean, and conclude that the samples come from populations with different means. It is this sort of inference that leads to the term “inferential statistics.”

Running a T-test

For independent two-sample t-tests, you must select one or more variables to be tested and one grouping variable. SPSS calculates group means, standard deviations, and sample sizes.

In our example, we will be testing the data to see if student math score (measured by MATHSCOR) or student computer score (measured by COMPSCOR) varies as a function of student math anxiety (measured by MANX). We will use MATHSCOR and COMPSCOR as the dependent variables to be tested, and MANX as the independent variable for grouping.

Where to Go From Here

The rest of this document has been intentionally deleted.

To find out how you may use the full version of this document and many other of our award-winning materials in your own training classroom, visit:

http://ittraining.iu.edu/ematerials/