

Advanced Data Analysis using industry accepted and widely popular statistical package

Dr. Md. Abdus Salam Akanda

Professor

Department of Statistics

Faculty of Science

University of Dhaka

Dhaka, Bangladesh

E-mail: akanda@du.ac.bd

Week#2: Introduction to Computer Applications on Statistical Analysis and Data Entry

■ Course Outline

- Functions of the application
 - Operating application
 - Different functionalities and how to perform
 - Types of files in the application
 - Creation of data files
 - Variable naming convention
 - An illustrative example
-
- **TEXT BOOK: M. A. Salam Akanda (2018). *RESEARCH METHODOLOGY -- A Complete Direction for Learners*, 2nd Edition, Akanda & Sons Publications, Dhaka, Bangladesh**



SPSS[®]
Real Stats. Real Easy.[™]

Why Have We Chosen to Work with SPSS?

- **Statistical Package for the Social Sciences (SPSS)** was chosen because of its popularity within both academic and business circles, making it the most widely used package of its type.
- SPSS is also a versatile package that allows many different types of analyses, data transformations, and forms of output - in short, it will more than adequately serve our purposes.
- SPSS is widely coveted due to its straightforward and English-like command language and impressively thorough user manual.

Users of SPSS

- SPSS is used by market researchers, health researchers, survey companies, government entities, education researchers, marketing organizations, data miners, and many more for the processing and analyzing of survey data.

What is SPSS?

- The statistical package SPSS was first devised in 1966. Actually SPSS was developed as an analysis program for social scientists. One of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions.
- The abbreviation SPSS stood for “Statistical Package for Social Sciences”.
- The SPSS Company gave the old abbreviation a new meaning (not very modest): Superior Performing Software System. One of the strong points of SPSS is that it can perform almost any statistical analysis.

Different Versions of SPSS

- SPSS – X is the SPSS version for minicomputers and mainframe computers.
- SPSS/PC+ is the SPSS version for DOS based computers
- **SPSS for Windows**
- SPSS – X can use up to 315 variables in comparison to the 500 that SPSS/PC+ can use.
- SPSS for Windows can use more than 500 variables.

Different Versions of SPSS (con't)

- SPSS for Windows has been derived from the mainframe version and not from SPSS/PC+ version.
- The exchange of files between the different versions of SPSS (SPSS – X, SPSS/PC+, SPSS for Windows) is handled by special SPSS files that are created and read with the **IMPORT** and **EXPORT** commands. Communication with other well-known PC packages is also possible.
- We will stick to **SPSS for windows**. SPSS for windows is an advanced statistical package designed to run interactively on PC and other computers in a graphical environment, using descriptive menus and simple dialog boxes to do most of the work. Most tasks can be accomplished simply by **pointing and clicking** the mouse.

What can SPSS do?

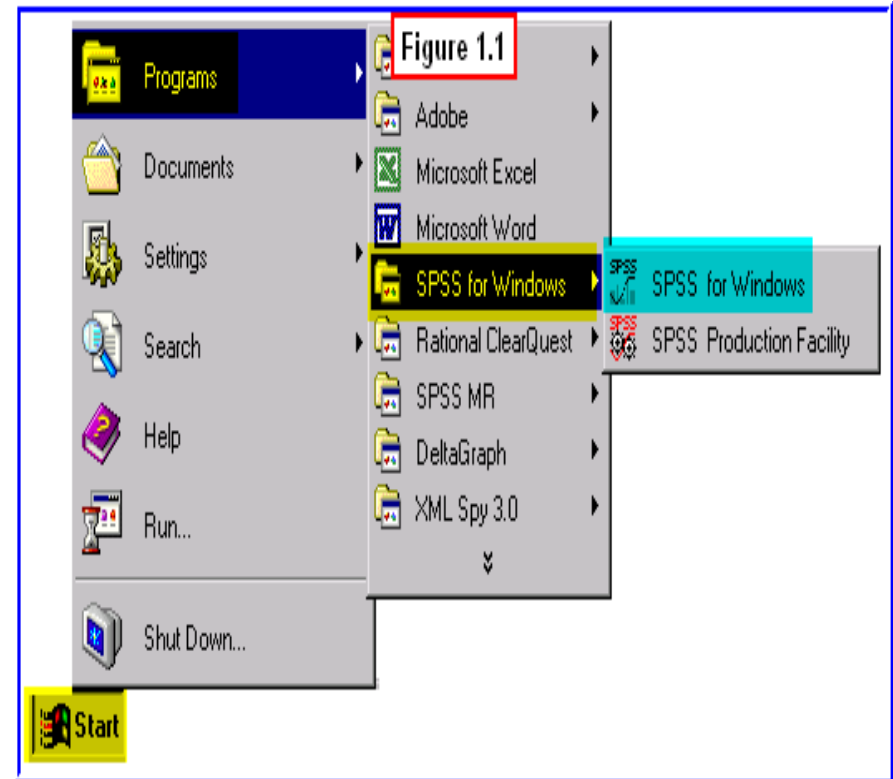
- Data entry
- Manipulate and manage data
- Produce reports and tables
- Produce graphical output
- Perform simple and complex statistical analyses

Core Statistical Functions of SPSS

- There are a handful of statistical methods that can be leveraged in SPSS, including:
 1. Descriptive statistics, including methodologies such as **frequencies**, **cross tabulation**, and **descriptive ratio statistics**.
 2. Bivariate statistics, including methodologies such as **analysis of variance (ANOVA)**, **means**, **correlation**, and **nonparametric tests**.
 3. Numeral outcome prediction such as **linear regression**.
 4. Prediction for identifying groups, including methodologies such as **cluster analysis** and **factor analysis**.

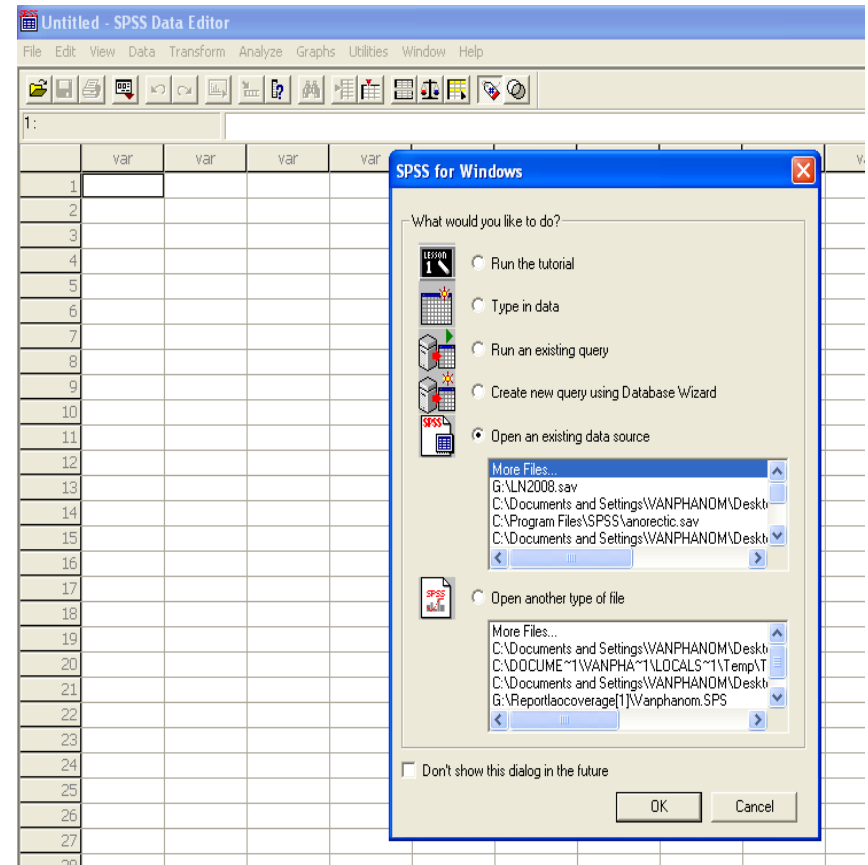
Starting SPSS for Windows

- You can start SPSS either by Using **Start menu** or by using **Shortcut icon**.
- **Using Start menu:** Click on Start menu at the bottom-left corner of your screen. Point the cursor to Programs, point the cursor to SPSS Inc. and then click on **SPSS Statistics 20.0**.
- **Using Shortcut icon:** Double click on Shortcut icon of SPSS Statistics 20.0 on the desktop.



SPSS for Windows

- There are two icons for SPSS (if you start it from start button)
- **SPSS Statistics 20.0 (This will be of our interest)**
- **SPSS Statistics 20.0 License Authorization Wizard.lnk**
- If we start SPSS Statistics 20.0, the following window will be open. If a dialog box appears, click on **Cancel** or press the **ESC** key from key board or click on **cross button** (✕) in the top-right corner. Then the dialog box will disappear. Now existing window is **SPSS Statistics Data Editor** window.

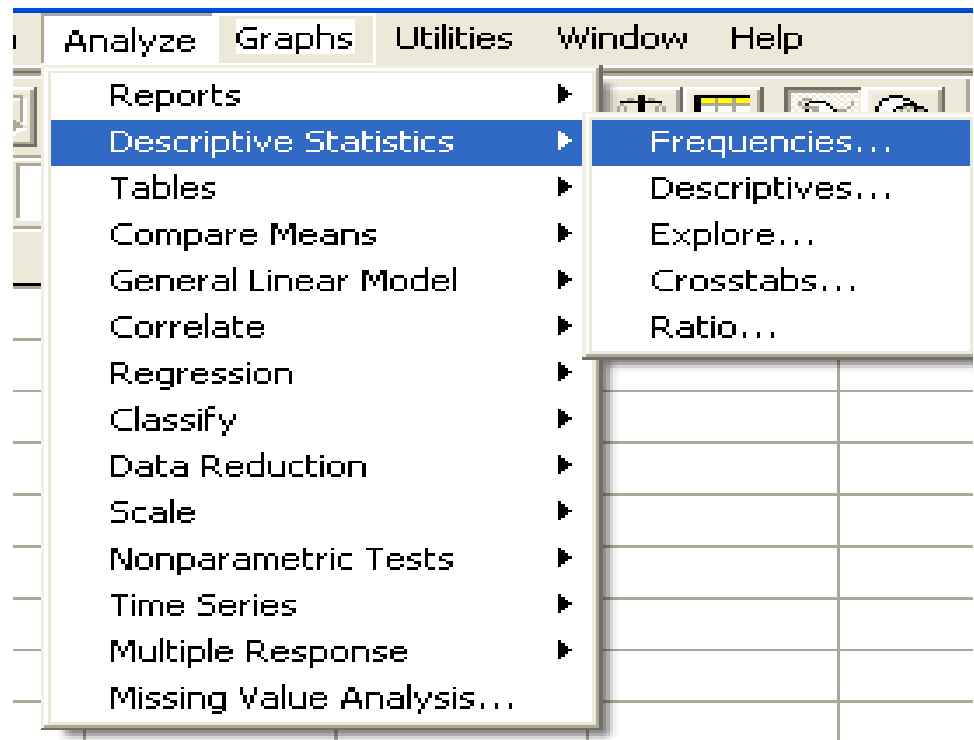


Menu Bar

Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

- The Menu bar lists 12 pull down menu, grouping the available SPSS commands. Some of these have sub-menus, the Analyze menu is like this.



The Toolbar



- The toolbar, located just below the menu bar, provides quick and easy access to many frequently used facilities.
- **Open File:** Displays the Open File dialog box for the type of windows that is active.
- **Save File:** Saves the working file, if the file has no name, it displays the Save File dialog box for the type of document that is active.
- **Print:** Displays the print dialog box.



Four-Windows in SPSS

- The four windows in SPSS are:
 1. Data Editor
 2. Output Viewer
 3. Chart Editor Window
 4. Syntax Window

Data Editor Window

- The Data Editor is the active window when you start SPSS. It is used to record all the data we want to analyze.
- It has two views: the **Variable View** and the **Data View**.
- The Variable View allows us to name each column in the Data table and specify what sort of values the column will contain.
- The Data View contains a table with a large number of cells in rows and columns. The table can be very large with only a small part of it visible, in which case use the scroll bars on the edges of the window to move round the table.
- We almost always enter data in the same way. Normally each row represents an individual case and each column represents a variable.

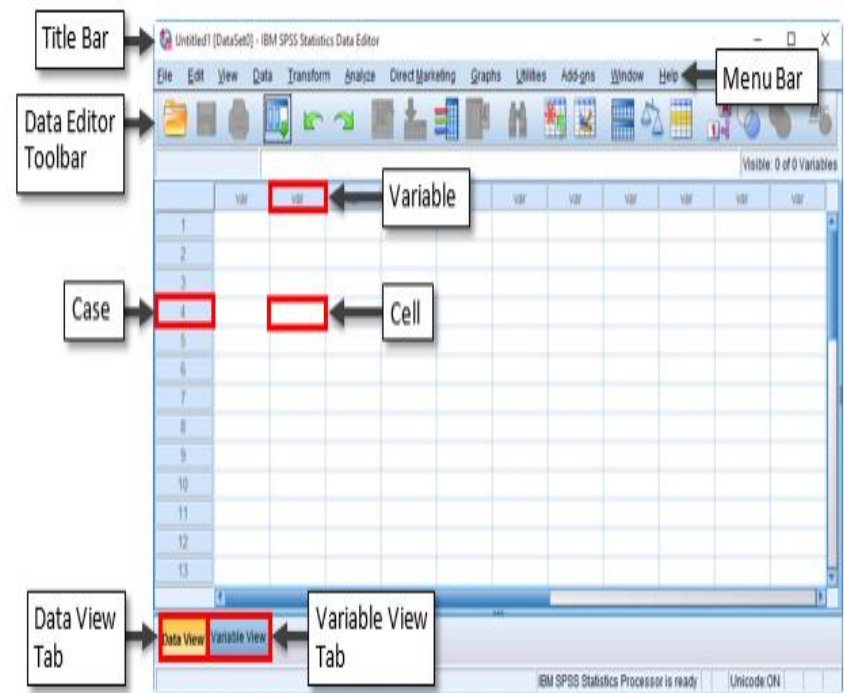
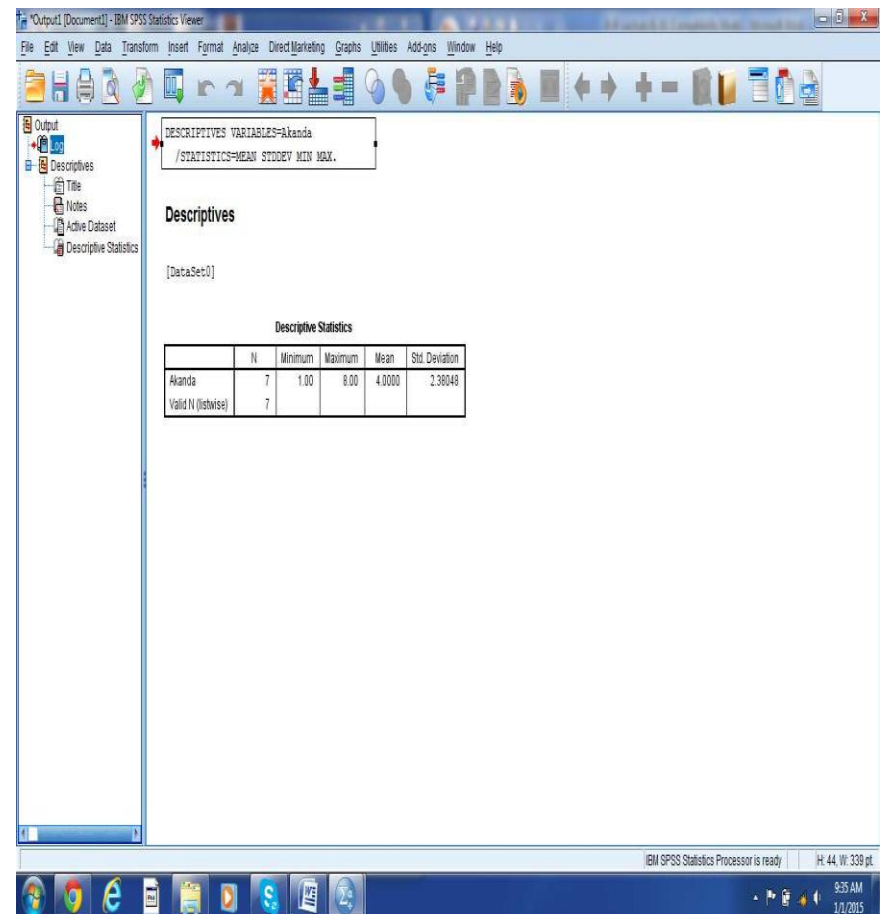


Figure 3 – IBM SPSS Statistics Data Editor Window

Output Viewer Window

- The viewer window is where we see the statistics and graphics – the output from the work in SPSS. The viewer window is also called Output Window which is split into two parts or panes:
 - The **Outline Pane** (Left side of the viewer window)
 - The **Display Pane** (Right side of the viewer window)



Pivot Table

- A *pivot table* is a table capable of dynamically displaying and rearranging multiple dimensions. There are three display areas of a pivot table: rows, columns, and layers. The dimensions of any table can be easily moved from row to column to layer. The ability to position multiple dimensions in each of these areas increases the number of data dimensions that can be viewed.

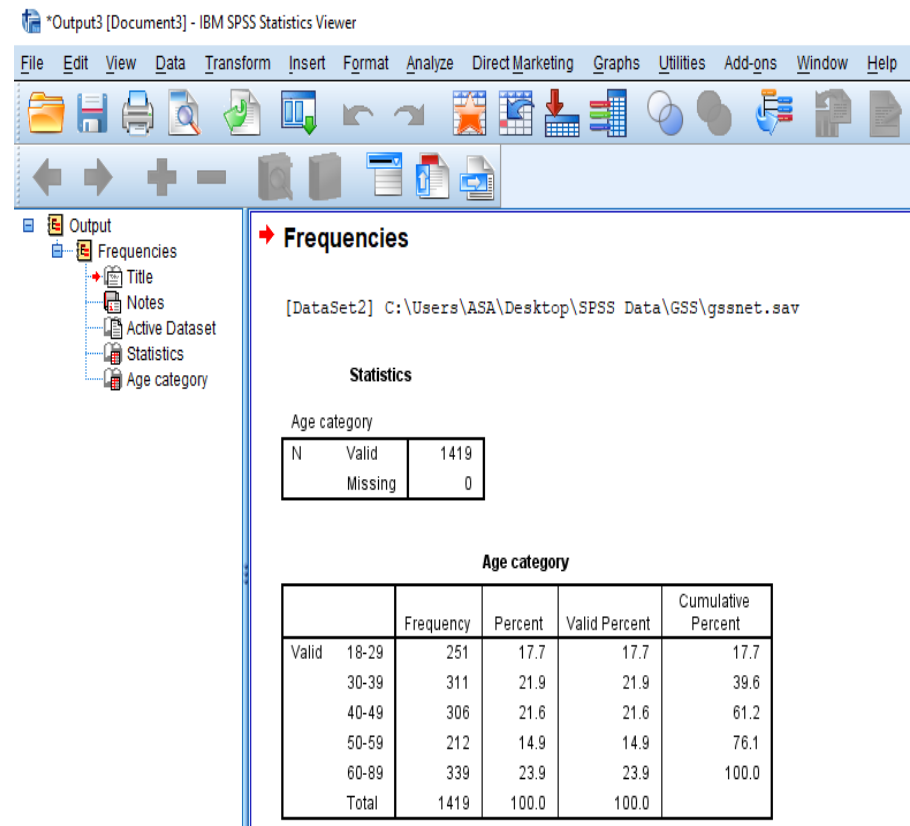
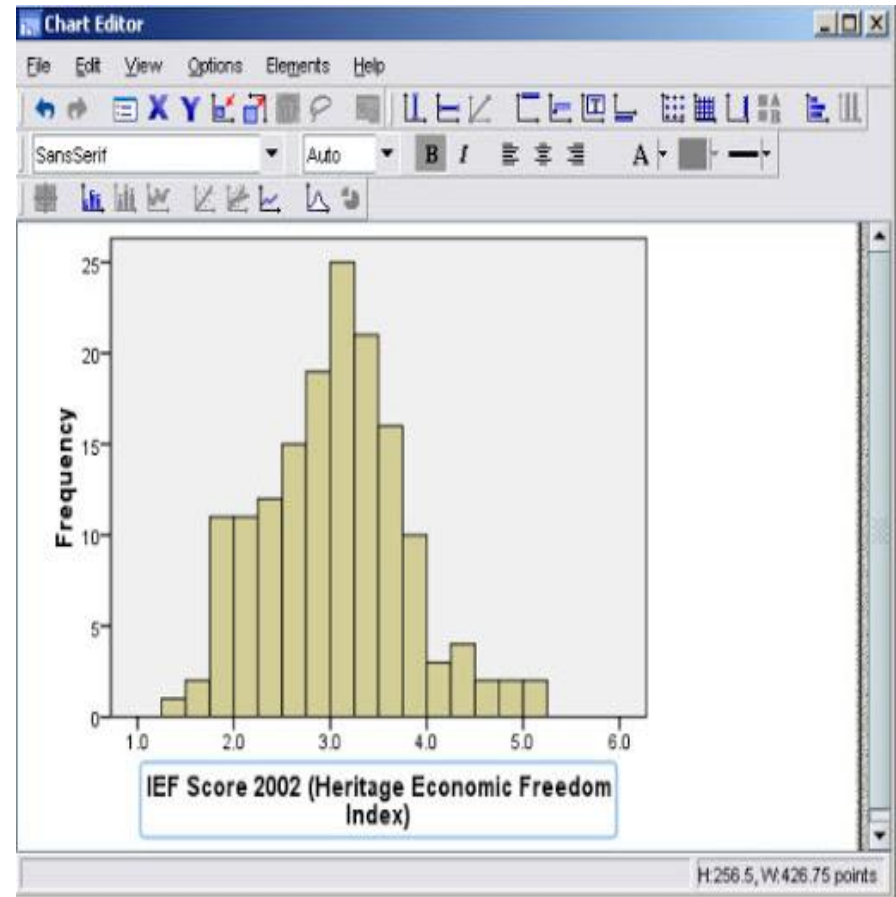


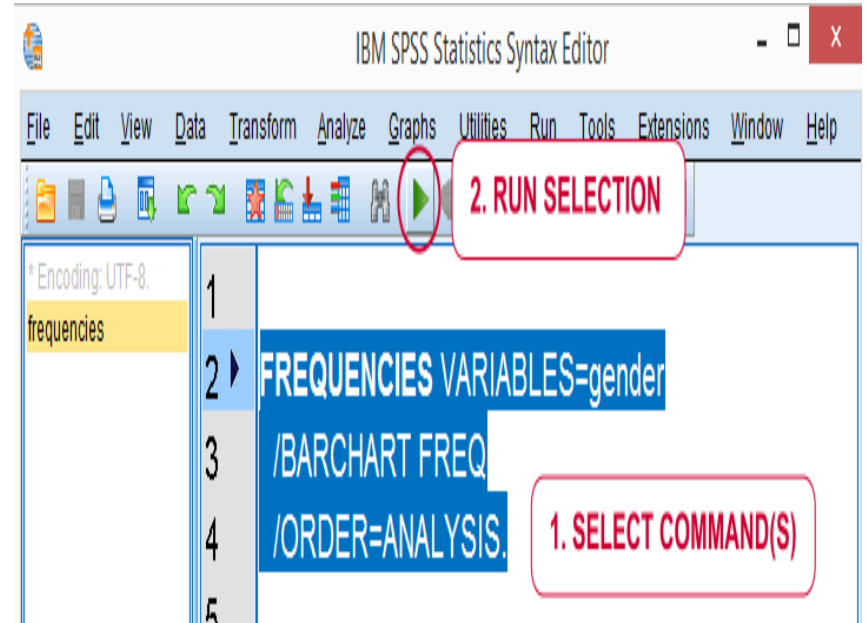
Chart Editor Window

- When you double-click a chart shown in an SPSS *Viewer* window you are opening the **SPSS Chart Editor**.
- The current chart can be modified using the various tools provided.
- Many of them are accessible from tool buttons or by changing the appropriate element in the **Properties dialog**. Double-clicking any chart element will open the **Properties dialog**; selecting a particular element on the chart shows the options available for it.



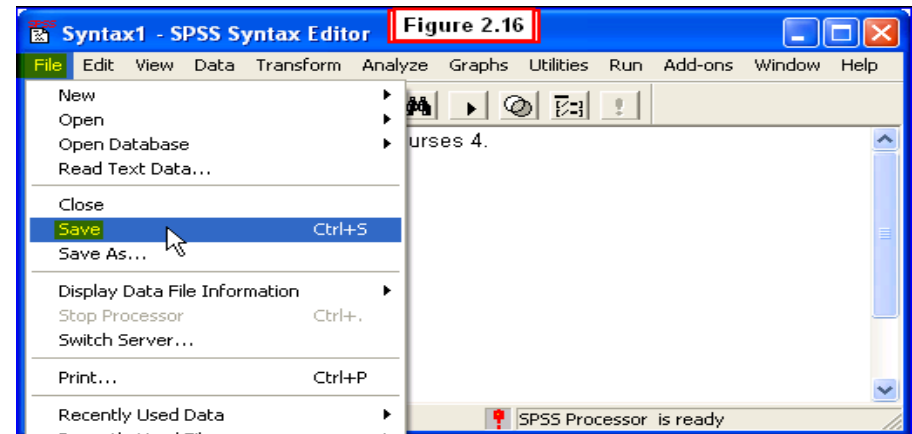
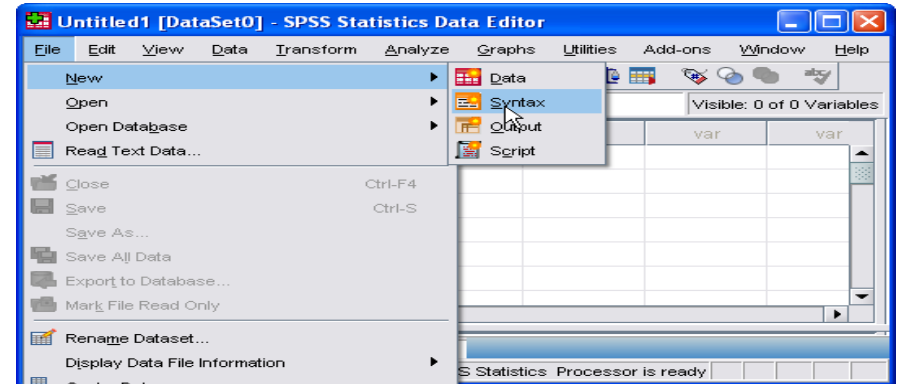
Syntax Window

- A syntax window is a window into which we can paste and/or write SPSS command. Then running the syntax on a data file we can get the desired output.



Create & Save a New Syntax File

- Click on **File** menu from the data window. Point the cursor to **New** and then click on **Syntax**. (A Syntax Window will appear).
- We can save the Syntax Window as well as Output Window at our desire location simply by pressing **Ctrl+S**.



Summary of the Four SPSS Windows

- **Data Editor** - a spreadsheet used to create data files and run analyses using menus.
- **Output Viewer** - a window displaying the results of analyses performed by SPSS.
- **Syntax Editor** - a text editor used to create files and run analyses using syntax code.
- **Chart Editor Window.** You can modify and save high-resolution charts and plots by invoking the Chart Editor for a certain chart (by double-clicking the chart) in a Viewer window. You can change the colors, select different type fonts or sizes, switch the horizontal and vertical axes, rotate 3-D scatterplots, and even change the chart type.

Various Types of Files in SPSS

- SPSS reads, creates and writes different types of files.
- Conventions for naming, printing, deleting or saving files and for submitting command files for processing differ from one computer to another or from one operating system to another.
- Each file is used to store a particular type of information.

Common Available Files in SPSS

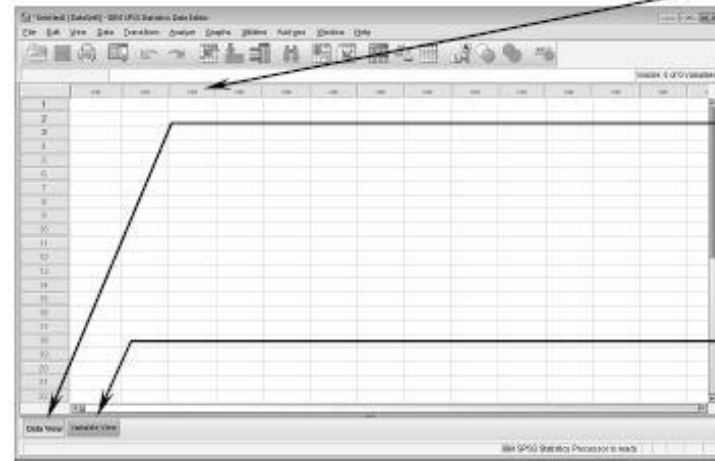
Type of File	Extension	Purpose
SPSS data file	SAV	Stores data along with the descriptive information of each variables and their values.
SPSS syntax file	SPS	Contains various commands and instructions to perform various task. Sometimes it contains inline data.
Output file	SPO	Contains outputs/ results generated after executing commands through menu or syntax.

Data Entry

- Data Entry in SPSS is the most important task involved in any analysis.
- You may create a data file using one of your favorite text editors, or word processing packages (e.g., Word Perfect, MS-Word). Files created using word processing software should be saved in text format before trying to read them into an SPSS session.
- You may enter your data into a spreadsheet (e.g., Lotus 123, Excel, dBASE) and read it directly into SPSS for Windows.
- Finally, you may enter the data directly into the spreadsheet-like *Data Editor of SPSS for Windows*.
- ***In this class we are going to examine one data entry methods: using the Data Editor of SPSS for Windows.***

Data View and Variable View

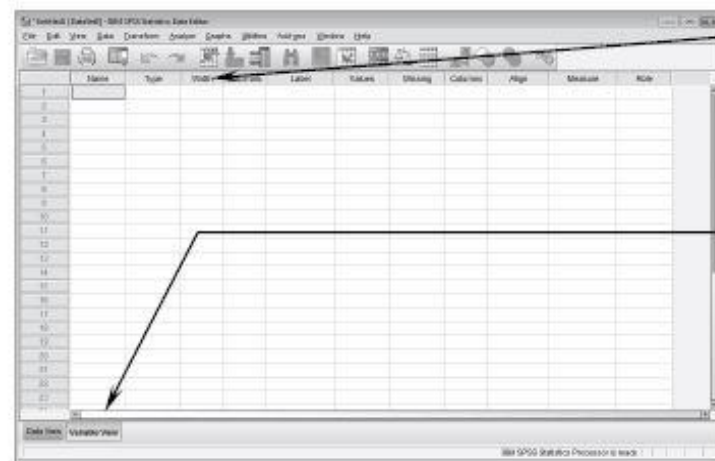
- The **Data View** is the screen you will use when entering your data into SPSS.
- The process of defining the variables is undertaken in the **Variable View**.



This is the Data View of the Data Editor window.

Note that the Data View tab is highlighted. This tells you that you are looking at the Data View.

Click on the Variable View tab to change to the variable view shown below.

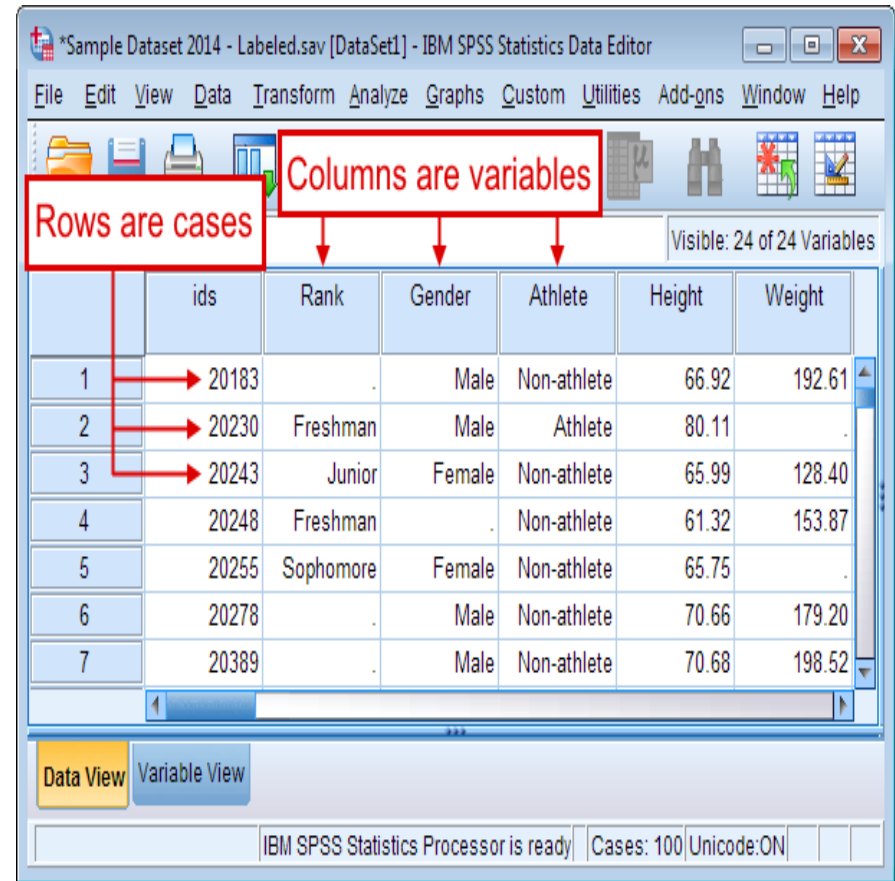


This is the Variable View of the Data Editor window.

Note that the Variable View tab is highlighted. This tells you that you are looking at the Variable View.

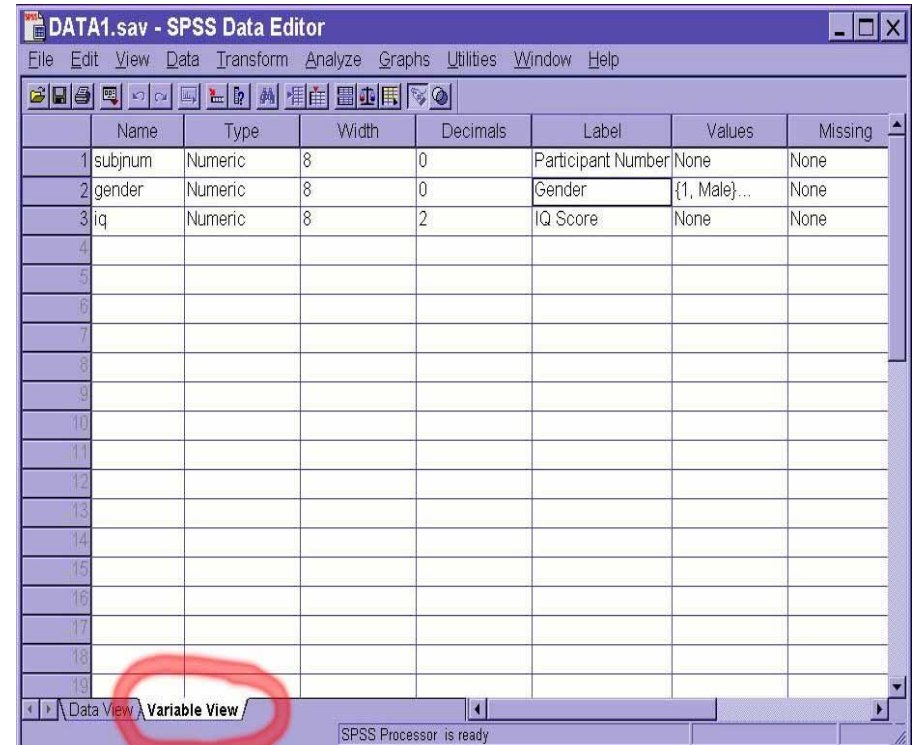
Variable & Case in SPSS

- A data set is the organized or structured form of data obtained from experiments, surveys or other sources. Before starting data analysis in SPSS, it is important to be clear about how the variables are recorded for each case/individual.
- Note that in the Data Editor window of SPSS each **column** will represent a **variable** and each **Row** will stand for a **case/respondent/individual**.



Setting up Variables

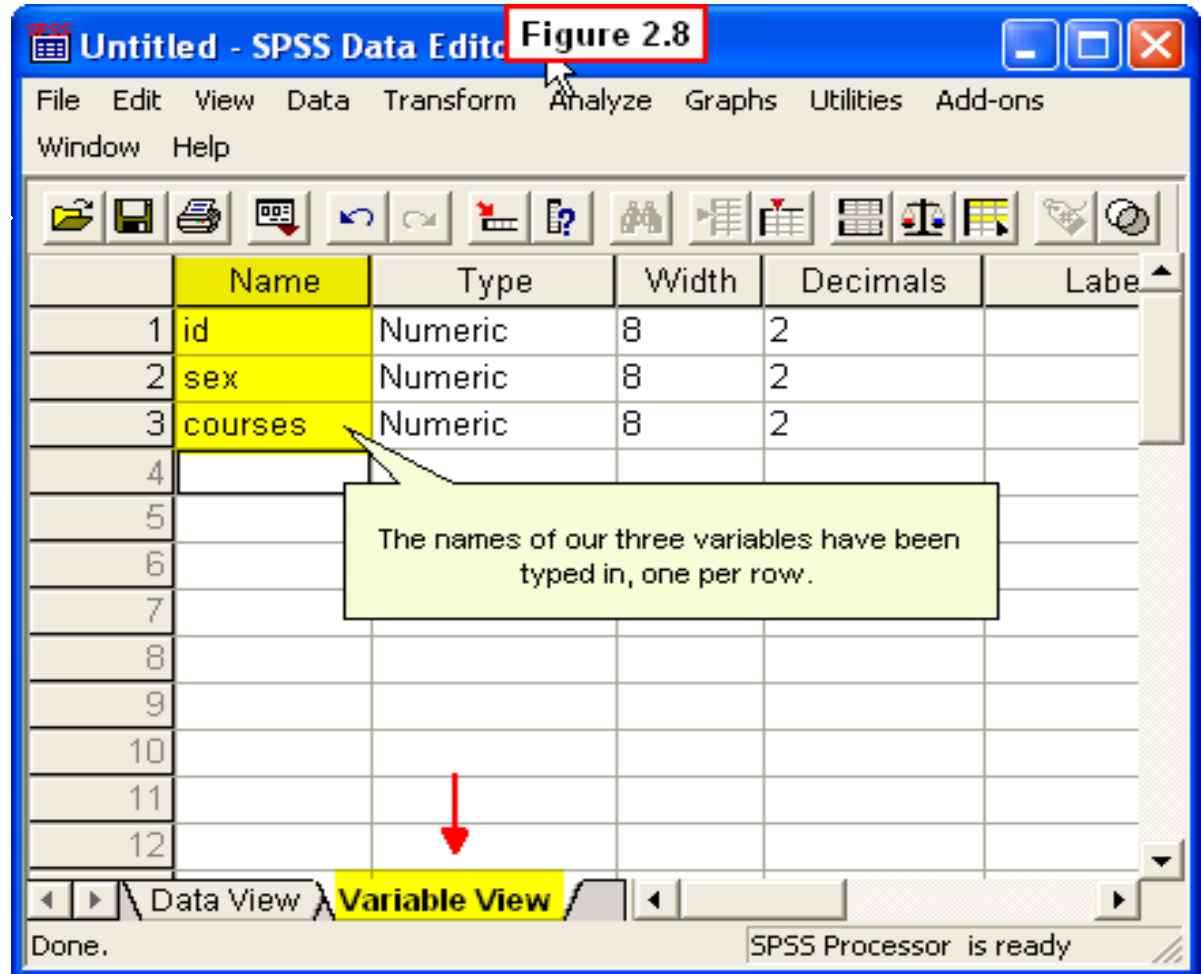
- If you are not already in the **Variable View** of the Data Editor, click on the **Variable View** tab now. We will now use this view to set up each of the variables we need.



Variable name

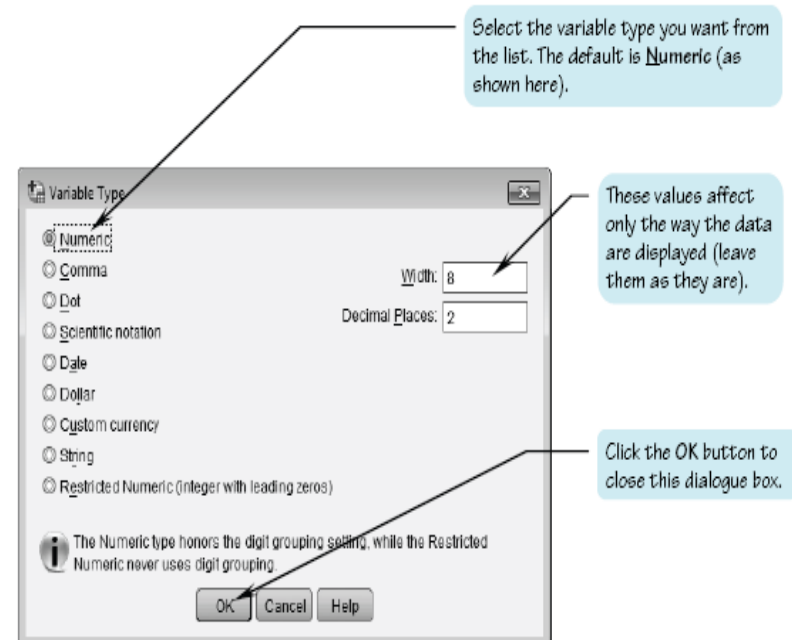
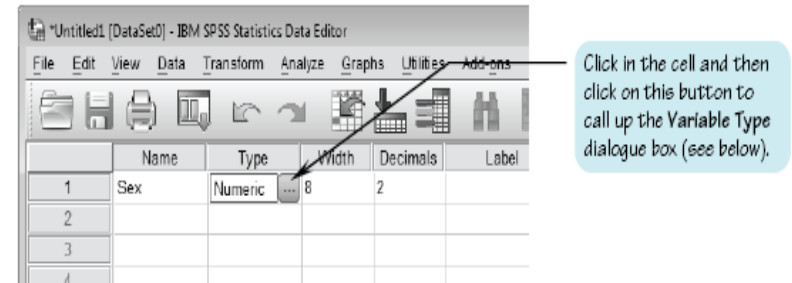
- Each variable name must be unique; duplication is not allowed.
- Variable names must start with a letter.
- Variable names may have up to 8 characters, including letters, numbers, and the symbols (full stops (.), @, #, _, or \$).
- Variable names cannot contain spaces or some special characters such as colons (:), semicolons (;), hyphens (-) or commas (,).
- Variable names cannot end with a period.
- Variable names that end with an underscore (_) should be avoided.
- The certain key words are reversed and may not be used as variable names, e.g. “compute”, “sum” and so forth.
- Ex. Subject_ID, but not “subject-ID”, and not “Subject ID”.

Variable Name (Con't.)



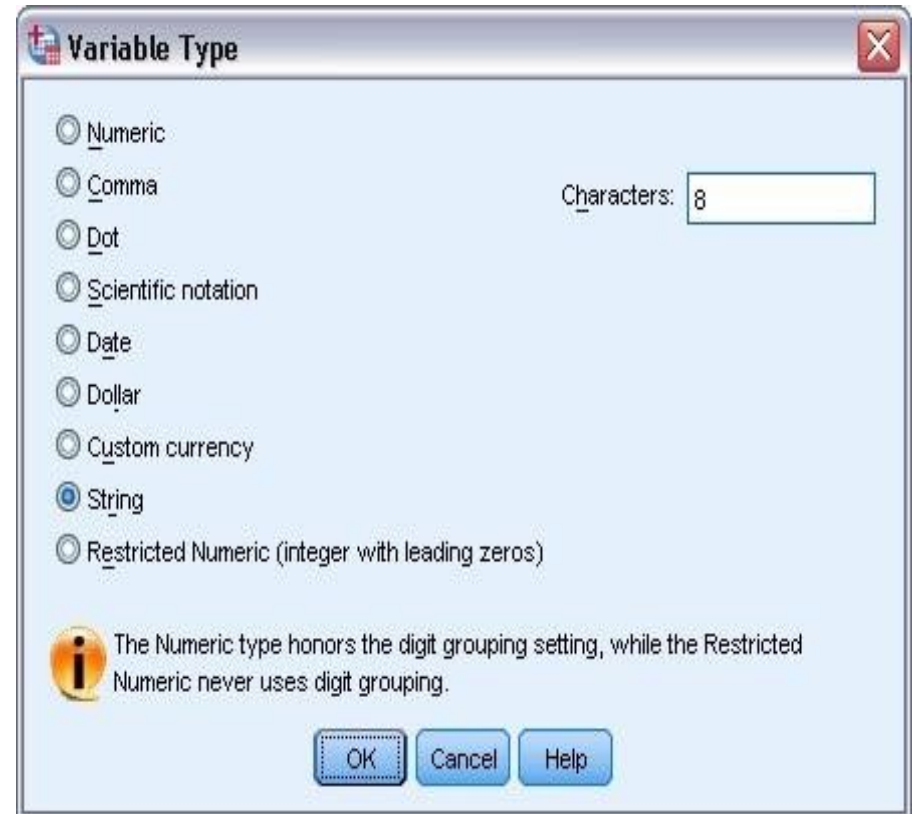
Variable type

- Basic type – numeric and string
- Maximum width for numeric variables is 40 characters, the maximum number of decimal positions is 16.
- String variables may contain letters or numbers. For string values a blank is considered a valid value.
- Numeric operations on the string variables will NOT be allowed, e.g. finding the mean, variance, standard deviation, etc...



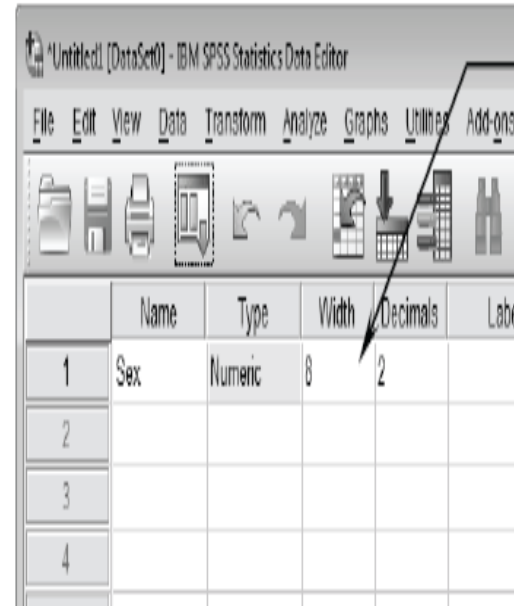
String Variable

- If you select a string variable, you can tell SPSS how much “room” to leave in memory for each value, indicating the number of characters to be allowed for data entry in this string variable.



Variable Width and Decimal Places

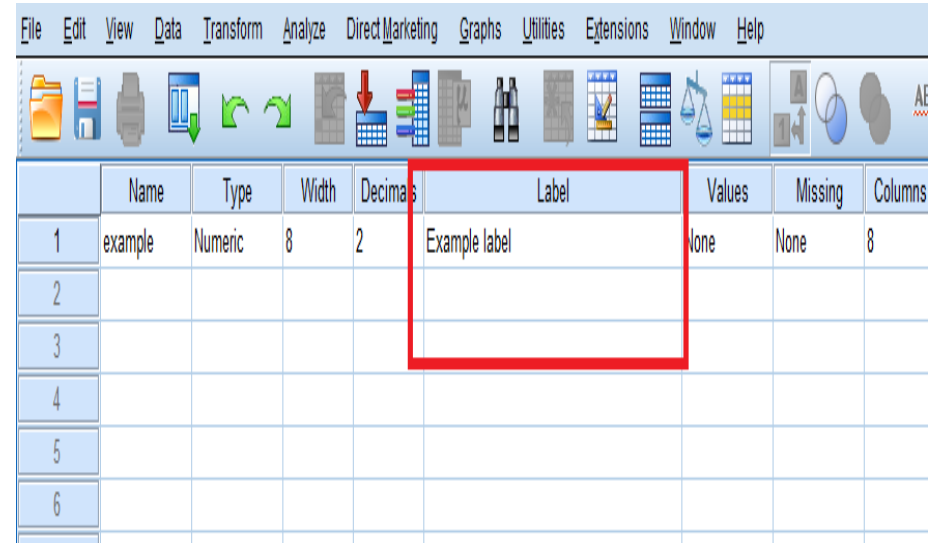
- **Width**
- The number of characters. SPSS will allow to be entered for the variable.
- For a numerical value with decimals, this total width has to include a spot for each decimal, as well as one for the decimal point.
- **Decimals**
- If more decimals have been entered or computed by SPSS, the additional information will be retained internally but not displayed on screen.



You can change the variable Width and number of Decimal places by changing these values. However, this only alters the look of the table.

Variable label

- A string to identify in detail what a variable represents.
- It is limited to 255 characters.
- It may contain spaces and punctuation.

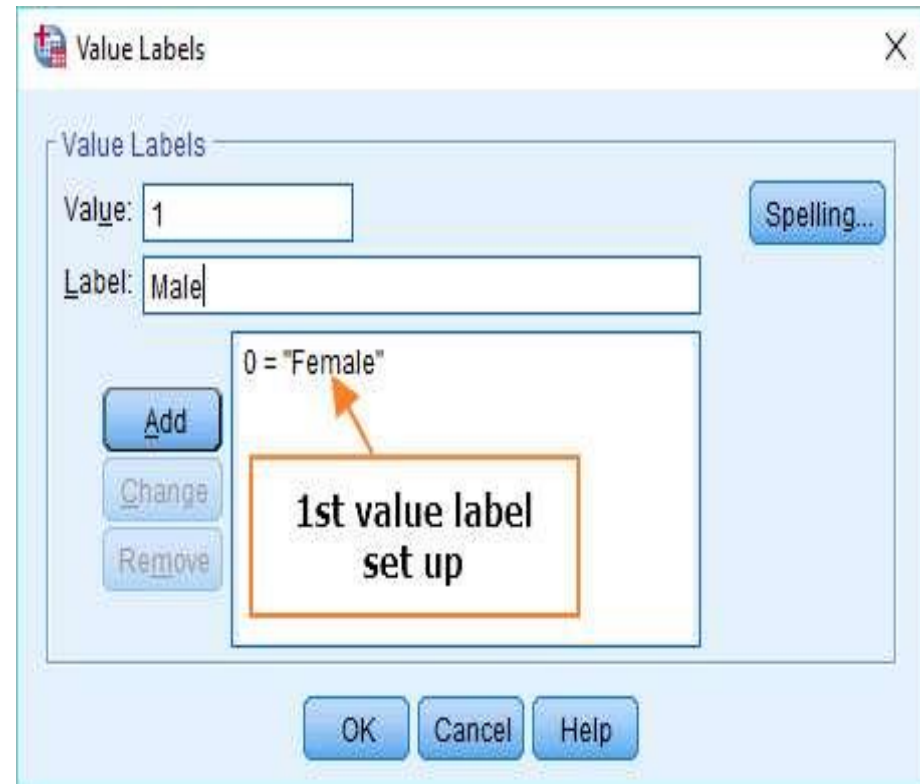


The screenshot shows the SPSS Variable View window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains icons for file operations, data manipulation, and analysis. The variable list table has columns: Name, Type, Width, Decimals, Label, Values, Missing, and Columns. The first variable, 'example', is of type 'Numeric' with a width of 8 and 2 decimal places. Its label is 'Example label', which is highlighted by a red rectangular box. The 'Values' column shows 'None', and the 'Columns' column shows '8'.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	example	Numeric	8	2	Example label	None	None	8
2								
3								
4								
5								
6								
-								

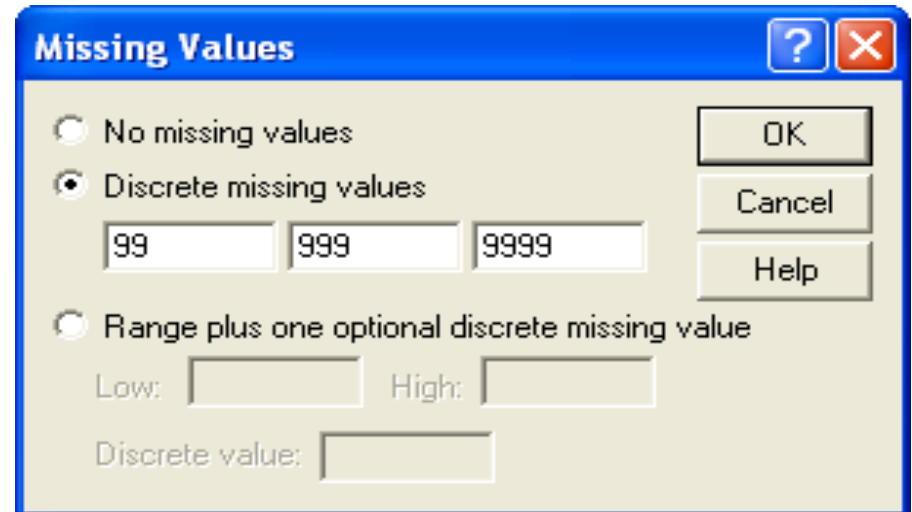
Value labels

- A value label is a label assigned to a particular value of a variable. You are most likely to use value labels for **nominal** or **categorical** variables.
- A second use for value labels is with a grouping or independent variable.



Missing Values

- Signal to SPSS which data should be treated as missing.
- System Missing data – SPSS display a single period.



The image shows the 'Missing Values' dialog box in SPSS. It has a blue title bar with a question mark icon and a close button. The dialog contains three radio button options: 'No missing values', 'Discrete missing values' (which is selected), and 'Range plus one optional discrete missing value'. Under 'Discrete missing values', there are three text boxes containing the values '99', '999', and '9999'. Under 'Range plus one optional discrete missing value', there are text boxes for 'Low:', 'High:', and 'Discrete value:'. On the right side, there are three buttons: 'OK', 'Cancel', and 'Help'.

Missing Values

☐ No missing values

☒ Discrete missing values

99 999 9999

☐ Range plus one optional discrete missing value

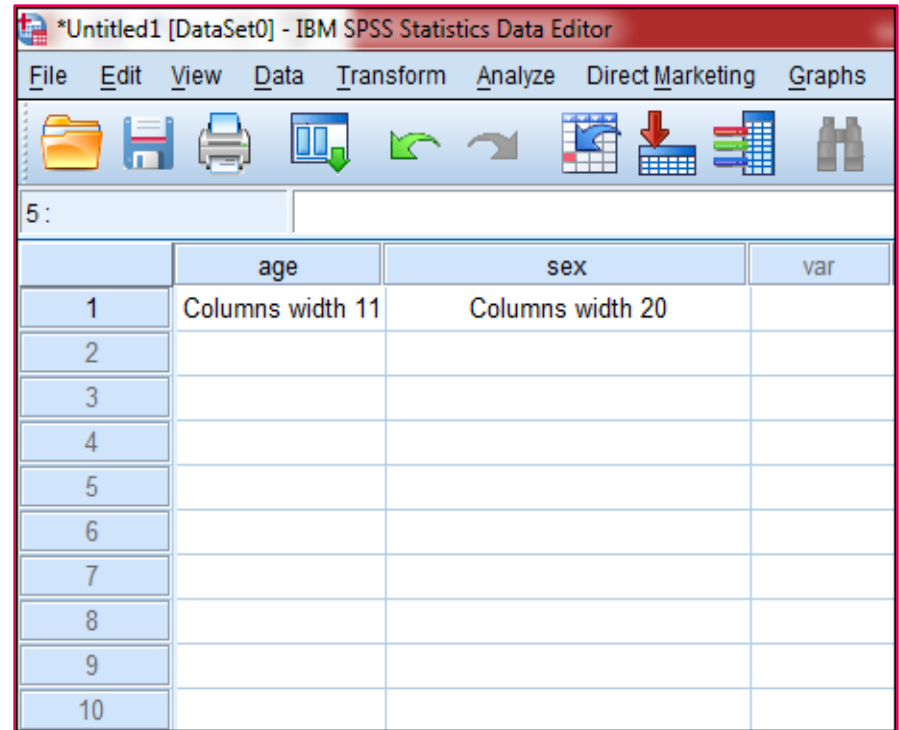
Low: High:

Discrete value:

OK Cancel Help

Column Format

- How wide the column should be for each variable
- Columns affect only the display of values in the Data Editor. Changing the column width does not change the defined width of a variable.







The screenshot shows the IBM SPSS Statistics Data Editor window titled '*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, and Graphs. The toolbar contains icons for file operations, data manipulation, and visualization. The data grid has 10 rows and 4 columns. The first column is labeled '5:'. The second column is labeled 'age' and has a width of 11. The third column is labeled 'sex' and has a width of 20. The fourth column is labeled 'var'.

5:	age	sex	var
1	Columns width 11	Columns width 20	
2			
3			
4			
5			
6			
7			
8			
9			
10			

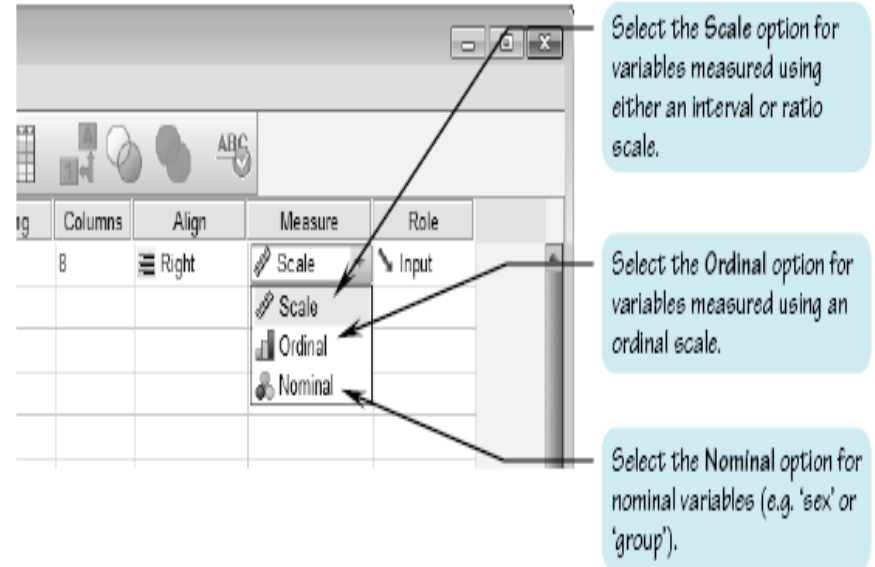
Column Alignment

- The column of the Variable View labeled **Align** allows you to specify the alignment of the text within the cells of the Data View table. This setting has no effect on the operation of SPSS and only changes the appearance of the Data View table.

Columns	Align	Measure
		
	 Left	
	 Right	
	 Center	

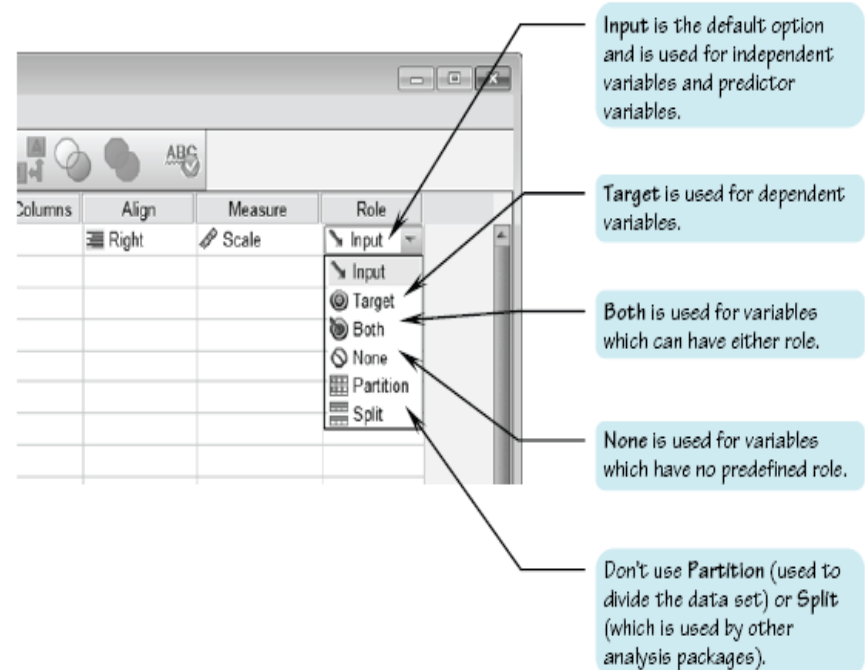
Measure

- The next column of the Variable View table is labeled **Measure**. This column is used to specify the level of measurement for the variable. SPSS offers three options: **Nominal**, **Ordinal** and **Scale**. Researchers usually distinguish four levels of measurement: nominal, ordinal, interval and ratio (see First class). SPSS does not distinguish between interval and ratio data and uses the term Scale to describe a variable measured using either of these levels of measurement.



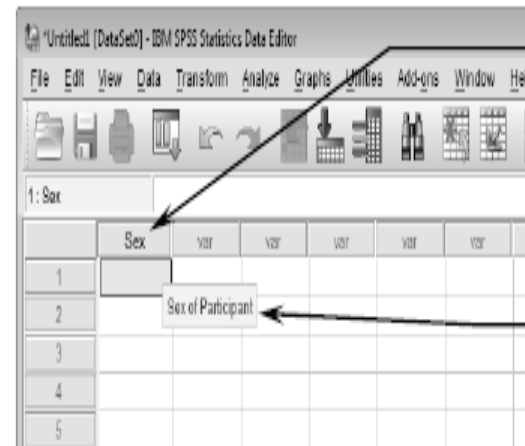
Role

- The final column of the Variable View table is called **Role**. This is a recent addition to SPSS, and is intended to help users who are undertaking complex analyses. The idea is that you can identify a group of variables as having a particular role in your analysis.



Entering More Variables

- Now switch back to the Variable View and repeat this process for each of the variables required for your data file.



The new variable name appears at the top of the column. This column is now ready to accept data.

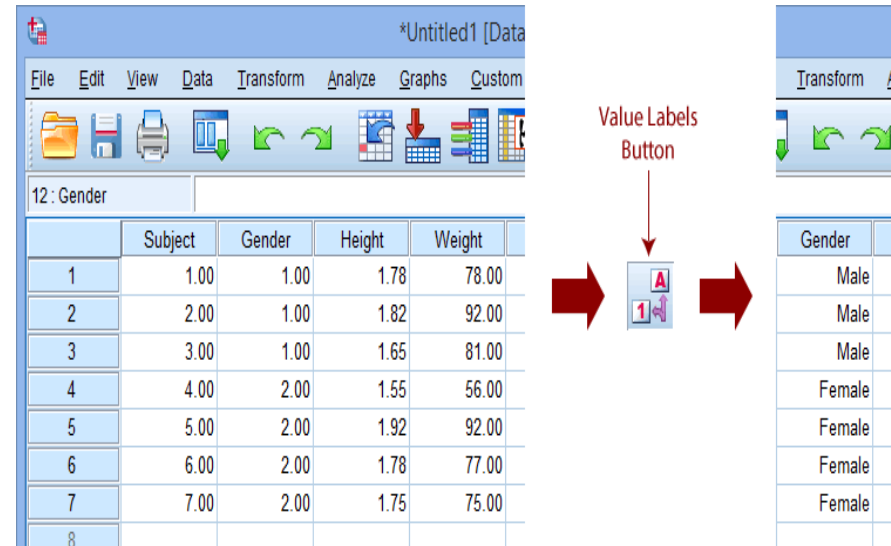
Note that if you move the mouse pointer over the variable name this pop-up displays the variable label.

Entering the Data

- The first step for entering the actual data is to click on the **Data View** tab.
- To enter new data, click in an empty cell in the first empty row. The “Tab” key will enter the value and jump to the next cell to the right. You may also use the Up, Down, Left, and Right arrow keys to enter values and move to another cell for data input.
- To edit existing data points (i.e., the change a specific data value), click in the cell, type in the new value, and press the Tab, Enter, Up, Down, Right, or Left arrow keys.

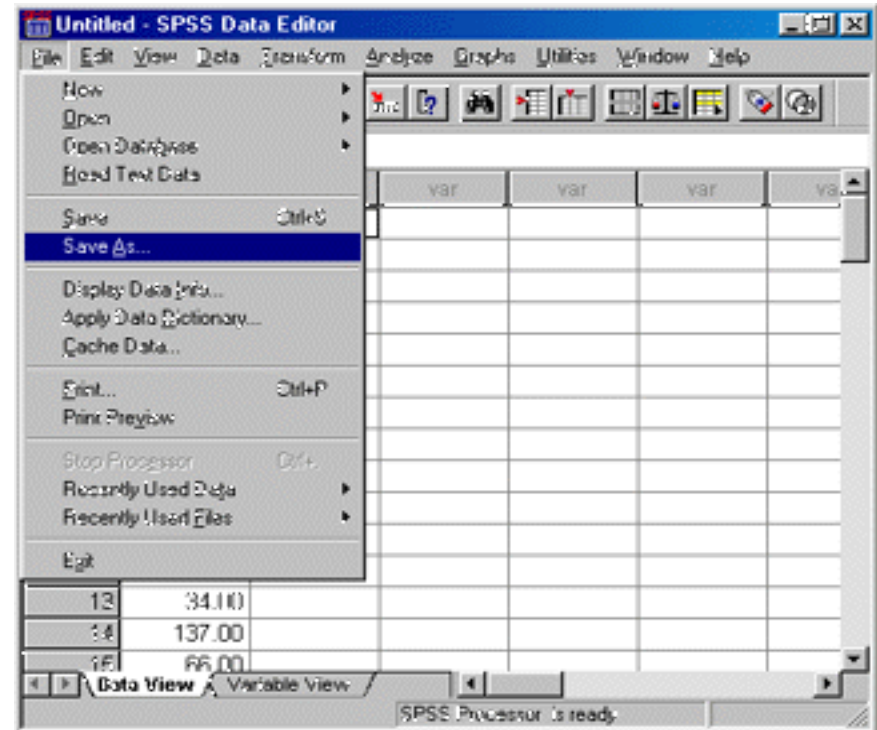
Value Labels Button

- If you have assigned value labels to one or more of your variables, you can choose whether you want SPSS to display the values you enter, or the labels associated with the values. For example, in this file, we have assigned the value labels 'Male' and 'Female' to the values 1 and 2 of the variable 'Sex'. SPSS can either display the values (i.e. the numerals '1' or '2') or the labels 'Male' or 'Female'. Clicking on the Value Labels button on the toolbar of the Data Editor window will toggle between these two display states (see picture). This option affects only the way the data are displayed in the Data Editor window, and not the way they are entered or analyzed.



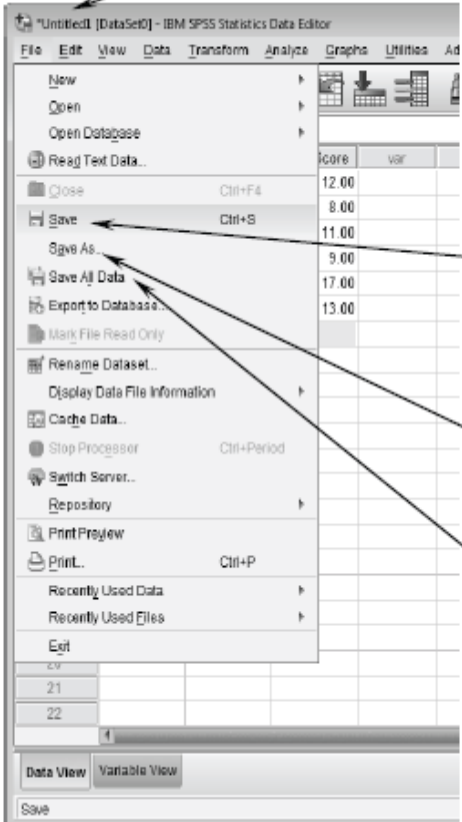
Saving a Data File

- You will have spent a lot of time entering your data, so remember to save the data file as soon as you have checked it carefully. If you are entering a large amount of data it is a good idea to save the file every few minutes.



To Save the Data to a File

- Click on the menu item File at the top of the screen. Now click on either **Save** or **Save As**.
- Select **Save** to resave the file using the existing name. The resaved file will replace the old version. If the file has not been saved previously, or if you click on **Save As**, you will be presented with the Save Data As dialogue box.



The screenshot shows the IBM SPSS Statistics Data Editor window. The 'File' menu is open, and the 'Save' and 'Save As' options are highlighted. The 'Save' option is associated with the keyboard shortcut 'Ctrl+S'. The 'Save As' option is associated with the keyboard shortcut 'Ctrl+Period'. The 'Save All Data' option is also visible. The background shows a data table with columns 'score' and 'var' and rows of numerical data.

SPSS uses the file name 'Untitled' for a file that has not been saved. Once you save the file your new name will appear here.

If there are two unsaved data files open at the same time, the second will be called 'Untitled2'.

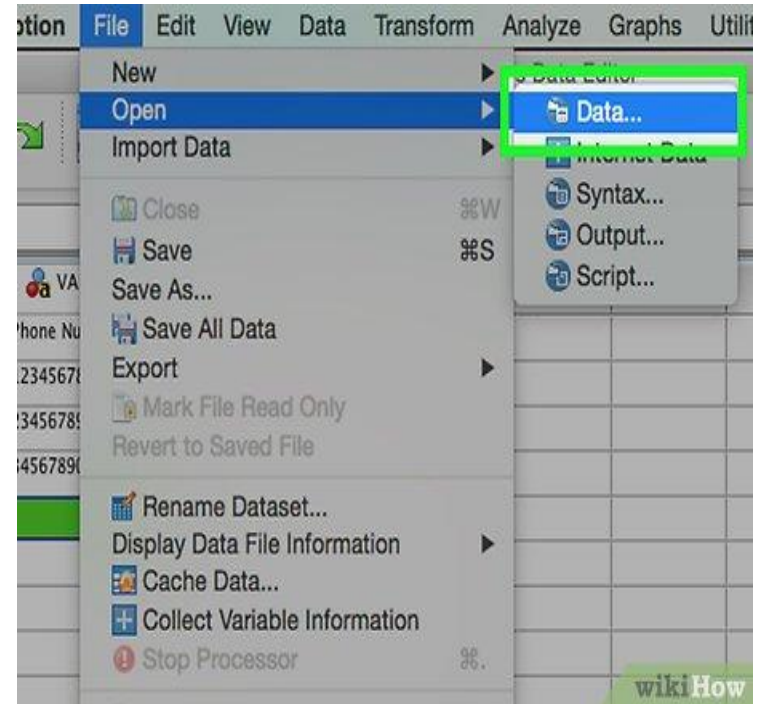
Click on **Save** to save the data file. If the file has been saved previously it will be re-saved using the same name. If not, you will be prompted for a name.

Click on **Save As** if the file has been saved before but you now want to save it under a different name.

Click on **Save All Data** if you have more than one file open, and you want to save them all.

Opening a Data File

- To open a data file follow the instructions below:
- Ensure that the Data Editor window is the active window. If this is not the case, select the SPSS icon from the taskbar at the bottom of the screen.
- To open a different data file, click on the **File** menu.
- Select **Open**.
- Select **Data**. The **Open Data** dialogue box will now appear.
- Select the appropriate folder from the drop down list in the box labelled **Look in**. The buttons next to this box can be used to move up one level in the directory, or to make a new folder, or to change the way the file information is displayed.
- Open the file either by selecting it from the list and clicking **Open**, or by double-clicking the file name.



Data Entry Exercise

Let us do create a data file in SPSS using the information given below:

ID	Name	Sex	Age	Region	Height	Education	Income
1	Ahmed Ali	Male	45	Dhaka	67	Higher	32000
2	John Abraham	Male	36	Chittagong	70	Secondary	14000
3	Meena	Female	23	Barisal	62	Secondary	13000
4	Ronjon Sharma	Male	42	Khulna	71	Illiterate	4000
5	Helal	Male	57	Rajshahi	65	Primary	9000
6	Nancy	Female	40	Dhaka	59	Higher	28000
7	Suzuka	Female	34	Rangpur	64	Secondary	11000
8	Mintu	Male	67	Rajshahi	57	Illiterate	7500
9	Romeo	Male	38	Khulna	68	Primary	9500
10	Anisul Haque	Male	41	Sylhet	69	Secondary	10000

Value labels: Sex: Male = 1, Female = 2

Education: Illiterate = 0, Primary = 1, Secondary = 2, Higher = 3

Summary

- In this class we described introduction to SPSS and importance of it in real life.
- We discussed different versions and important components of SPSS.
- We explained the different parts of the data window and showed how to define a variable.
- We have shown you how to create a data file in SPSS.
- We walked you through the process of setting up a data file, and showed how to save and open a data file.

**Thank you so much
for hearing me.**

