" A Number speaks more than a thousand words."

In this modern world, business, education and all field of science have heavily rely on computer. This dependency has becomes so great that we are nearly impossible to understand the social and physical science problem and its associate research without substantial knowledge of statistics and some rudimentary understanding of statistical software. The number of statistical software packages are available and continue to grow each year. SPSS (Statistical Package for Social Sciences) is one of the popular package in academic and business circles. SPSS is a versatile package that allows many different types of analysis, data transformations, and forms of output. Before starting the concept of SPSS and its use we should know the basics of data and its type.

What is statistics?

'Statistics' refers to what is called statistical methods. It deals with the collection of data, their classification, analysis and interpretations of statistical data. Therefore, it is described as a branch of science which deals with classification, tabulation and analysis of numerical facts and makes decisions as well. Every statistical inquiry should pass through these stages.

If the data were collected from a random sample drawn from a well-defined population in such a way that every unit in the population has a known non-zero probability of being included in the sample, then the information derived from such sample can be generalized to the population (inferential statistics). If the data were collected from a non-random sample, then the information derived from the sample

cannot be generalized (descriptive statistics).

If data and variables are not properly organized in a computer, then computer software fails to provide meaningful results.

Data

Data is a set of collection of objects or cases. Each case has one or more attributes or qualities called variables. The word variable emphasizes that it is differences or variation. An element which changes or varies is called variable. It is a character under study. The values of variable vary from one to another. E.g. height of persons, marks secured by students in an exam, religion of people etc. Generally variable is classified into two types

i.   Qualitative variable

ii.   Quantitative variable

Qualitative Variable

The variable which varies in kind rather than in magnitude is called qualitative variable. It can be divided into different categories. It is also called categorical variable. Qualitative variables are present in nominal and ordinal scales. Nominal scale is simply a system of assigning number or the symbols to objects or events to distinguish one from another or in order to label them, whereas, in ordinal scale the quantification of items will be done by ranking. In this scale, the numerals are arranged in some order but the gaps between the positions of the numerals are not made equal. E.g. hair color, eye color, gender, smoking habit, academic achievement, social status etc.

 If the categories are meaningfully ordered, the variable is ordinal; if it doesn't matter in which way the categories are ordered, then the variable is nominal.

Quantitative Variable

The variable which varies in magnitude and can be expressed numerically is called quantitative variable. Quantitative variables are present in interval and ration scales. e.g. age, height, weight, income etc. Quantitative variable can be further divided into two types

   a. Discrete variable
   b. Continuous variable

Discrete Variable

A variable which can take whole number or countable number of values is called discrete variable. E.g. number of rooms in building, number of road accidents, family size, household size etc.

Continuous Variable

A variable which can take all possible values (whole number and fractions) within certain range is called continuous variable. e.g. height , weight, temperature, age etc.

Researcher need to be aware of the data they have since this affects what analyses are/not suitable.

Test of Hypothesis to be used
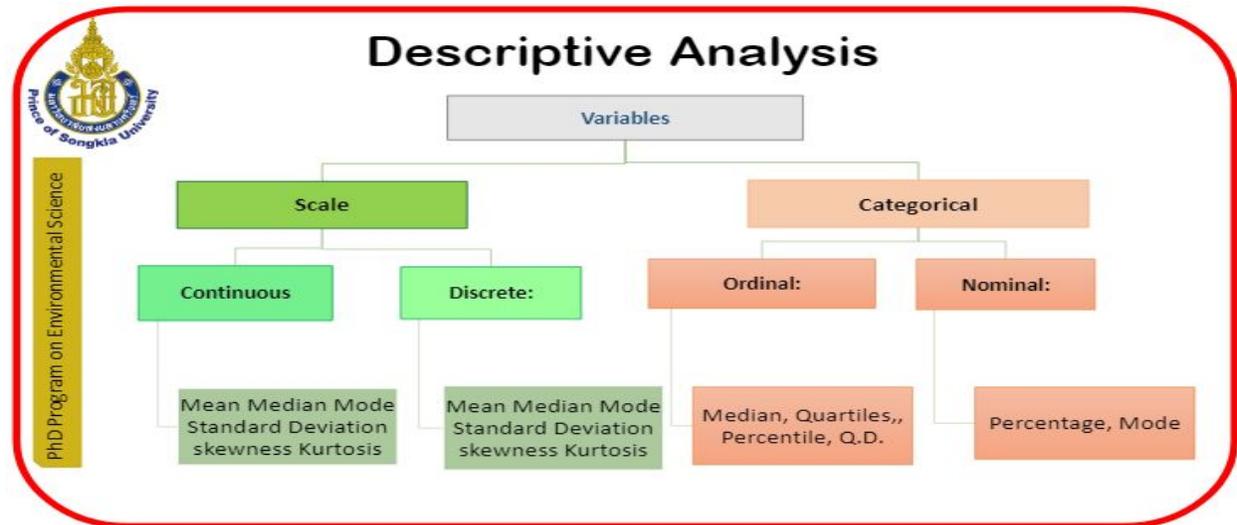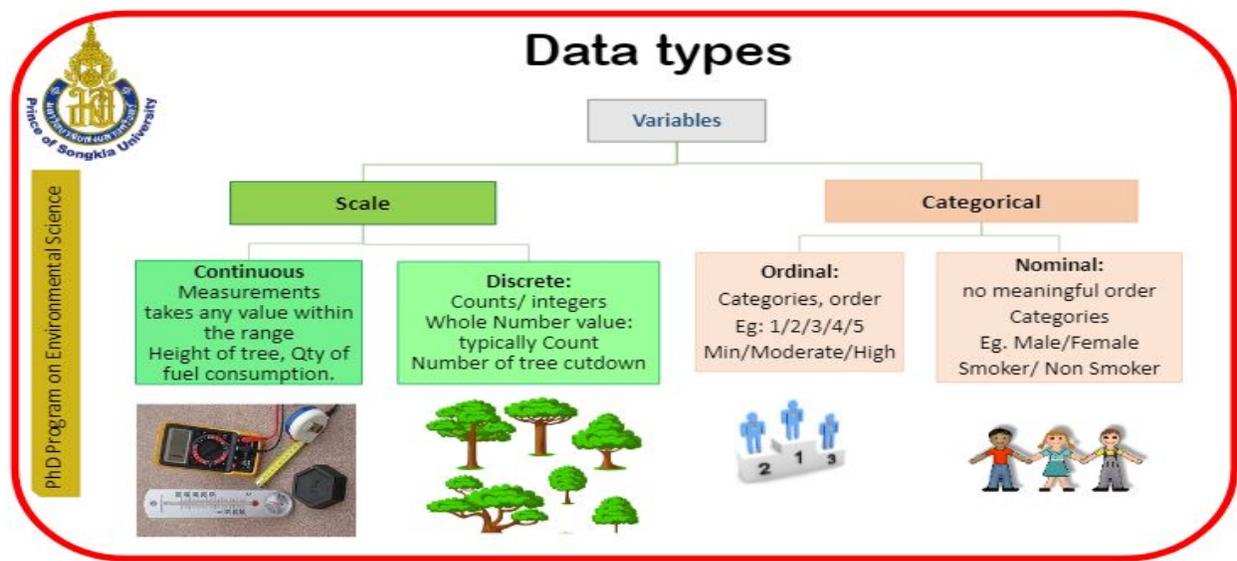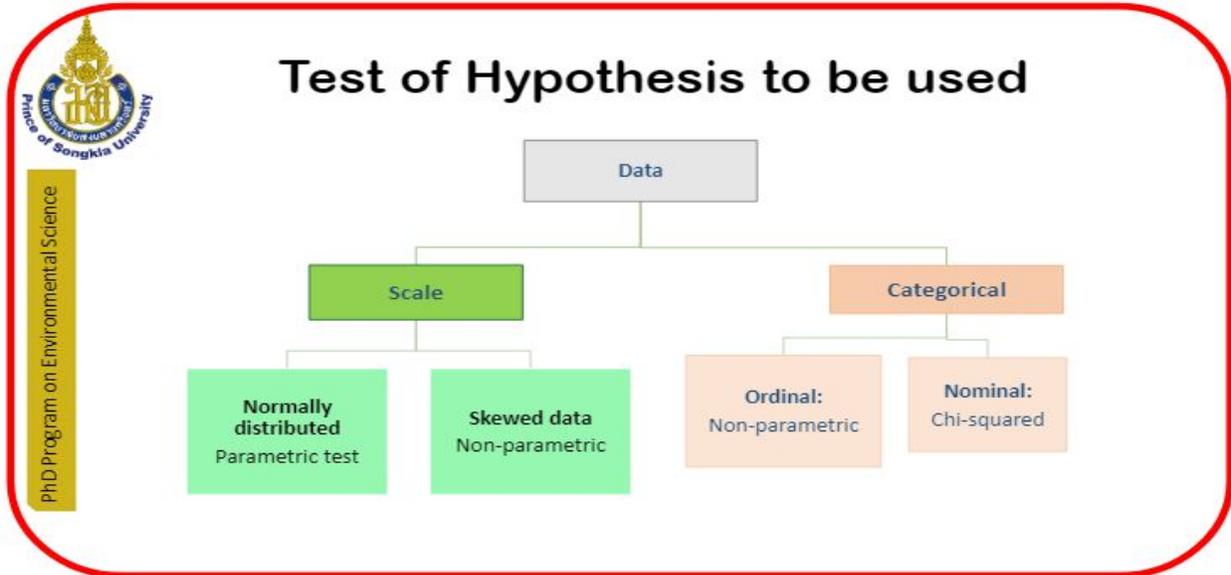
Population and Sample

A data frame is a collection, but a collection of what? Two important statistical terms are "population" and "sample." A population is the set of all the possible objects or units which might have been included in the collection. The root of the word "population" refers to people, and often one works with data frames in which the cases are indeed individual people. The statistical concept is broader; one might have a population of houses, a population of events such as earthquakes or coin flips. A sample is a selection of cases from the population. The sample size is the number of cases in the sample.

A **census** is a sample that contains the entire population. The most familiar sort of census is the kind to count the people living in a country. Nepal has a census every ten years.

Almost always, the sample is just a small fraction of the population. There are good reasons for this. It can be expensive or damaging to take a sample: Sample should be genuinely representative of the population. (In some fields, the ability to draw conclusions from a sample that can be generalized is referred to as external validity or transferability.) The process by which the sample is taken is important because it controls what sorts of conclusions can legitimately be drawn from the sample. One of the most important ideas of statistics is that a sample will be representative of the population if the sample is collected at random. In a simple

random sample, each member of the population is equally likely to be included in the sample. Ironically, taking a random sample, even from a single spot on the beach, requires organization and planning. If We want a sample of students at a particular university, We can get a list of all the students from the university registrar and use a computer to pick randomly from the list. Such a list of the entire set of possible cases is called a **sampling frame.** In a sense, the sampling frame is the definition of the population for the purpose of drawing conclusions from the sample. For instance, a researcher studying cancer treatments might take the sampling frame to be the list of all the patients who visit a particular clinic in a specified month. A random sample from that sampling frame can reasonably be assumed to represent that particular population, but not necessarily the population of all cancer patients. We should always be careful to define our sampling frame precisely. If We decide to sample university students by picking randomly from those who enter the front door of the library, We will get a sample that might not be typical for all university students. There's nothing wrong with using the library students for our sample, but We need to be aware that our sample will be representative of just the library students, not necessarily all students.It is wise to use a computer random number generator to select the cases to include in the sample.
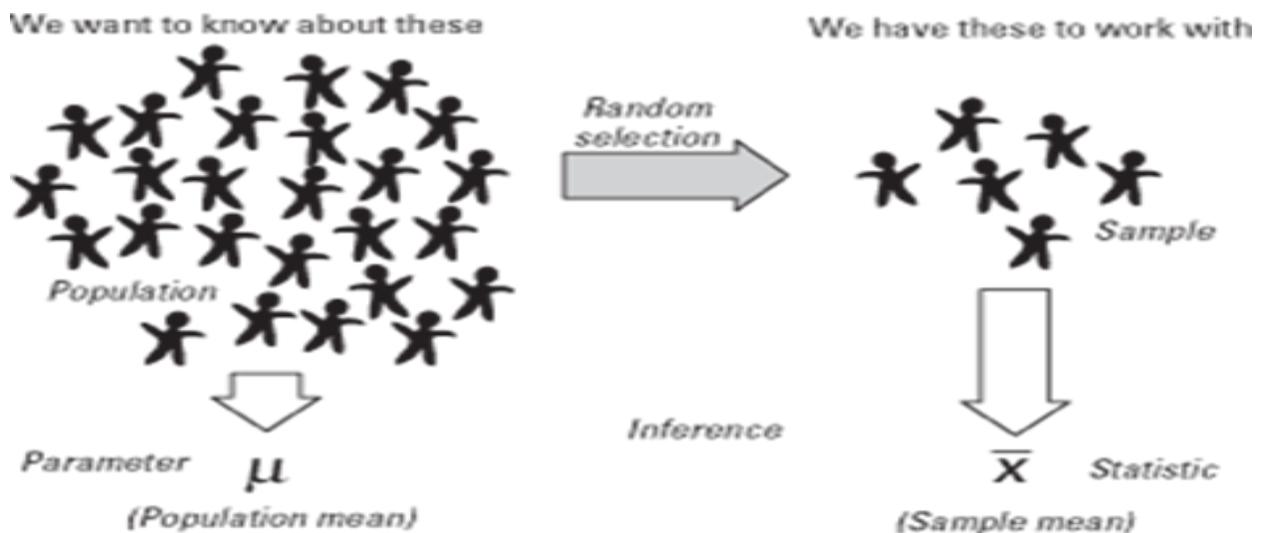
The main purpose of drawing sample is to draw conclusions about the population on the basis of sample data. The statistical tools of sample (called statistic) is used to estimate the value of population statistics (called parameter).

SPSS

SPSS package enable us to obtain statistics ranging from simple descriptive statistics to complex multivariate data analysis. We can plot the data in histogram, bar diagram, scatterplots and other ways. We can combine tow or more files, split files and sort items. We can modify existing variables and create new ones. In short we can do just about anything we have ever want in statistics with a set of data using this SPSS software package. In this module we will use SPSS 25.0 which I think that you have installed the package in your system.

Opening SPSS

If you have installed SPSS in your system then, click start, programs (or all programs), IBM SPSS Statistics, Σ IBM SPSS Statistics 25 (as in figure)
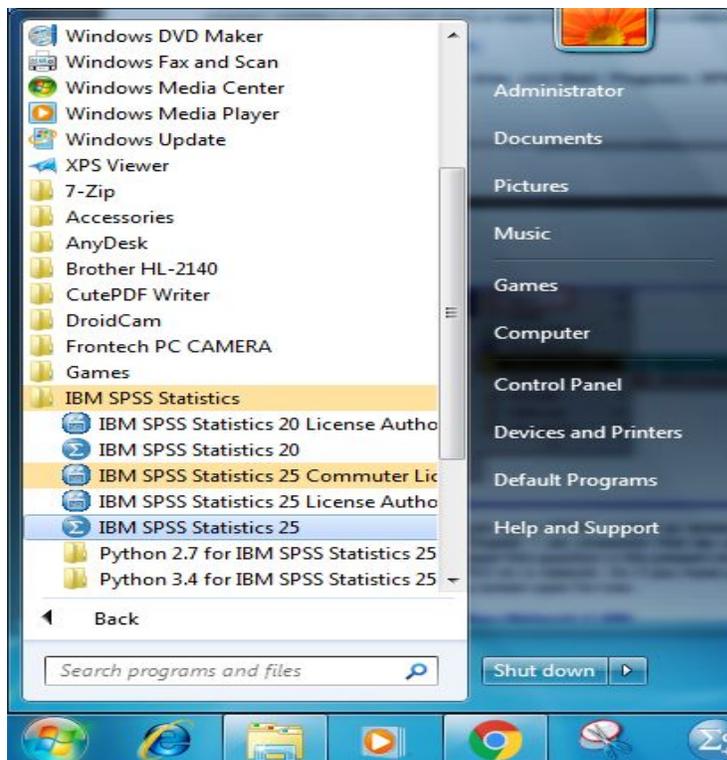
If you have spss file with extension *.sav then on clicking this file you can open SPSS software. You can either keep shortcut to your desktop or pinned the software in your taskbar.

Welcome window

This window allows the user to choose from an number of quick start options.New dataset, recent files, sample files, Help menu etc.
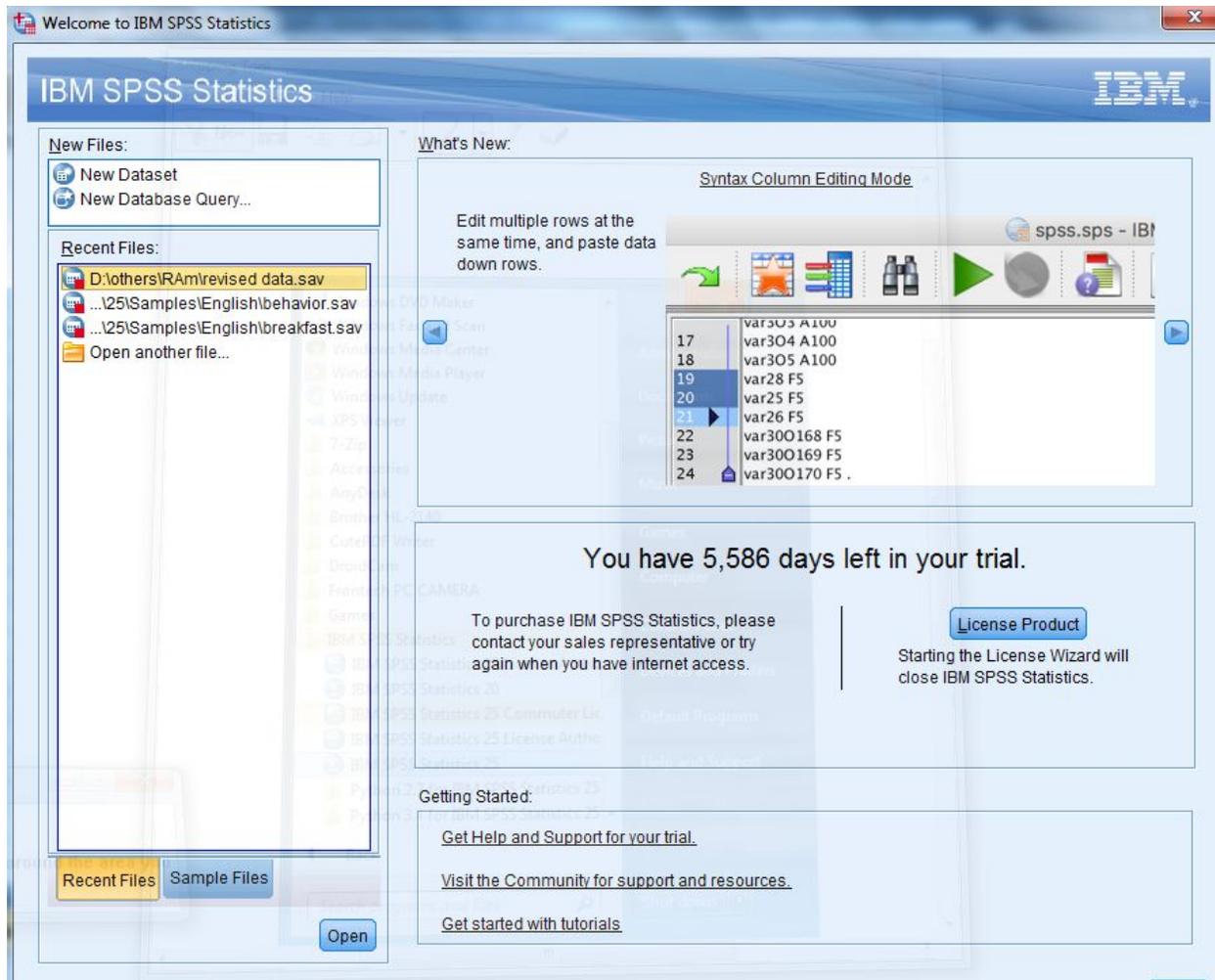
Different windows views in SPSS

There are a number of different types of windows in IBM® SPSS® Statistics:

Data Editor. The Data Editor displays the contents of the data file. we can create new data files or modify existing data files with the Data Editor. If we have more than one data file open, there is a separate Data Editor window for each data file. In data editor there are two views Data views and variable views

Two Views of Data Editor

Variable Views

Data Views

Output Viewer.

All statistical results, tables, and charts are displayed in the Viewer. We can edit the output and save it for later use. A Viewer window opens automatically the first time we run a procedure that generates output.

Pivot Table Editor.

Output that is displayed in pivot tables can be modified in many ways with the Pivot Table Editor. We can edit text, swap data in rows and columns, add color, create multidimensional tables, and selectively hide and show results.

Chart Editor.

As of the pivot table editor, we can modify high-resolution charts and plots in chart windows. We 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.

Text Output Editor.

Text output that is not displayed in pivot tables can be modified with the Text Output Editor. We can edit the output and change font characteristics (type, style, color, size).

Syntax Editor.

We can paste our dialog box choices into a syntax window, where our selections appear in the form of command syntax. We can then edit the command syntax to use special features that are not available through dialog boxes. We can save these commands in a file for further use. In our course we rarely use this syntax editor.
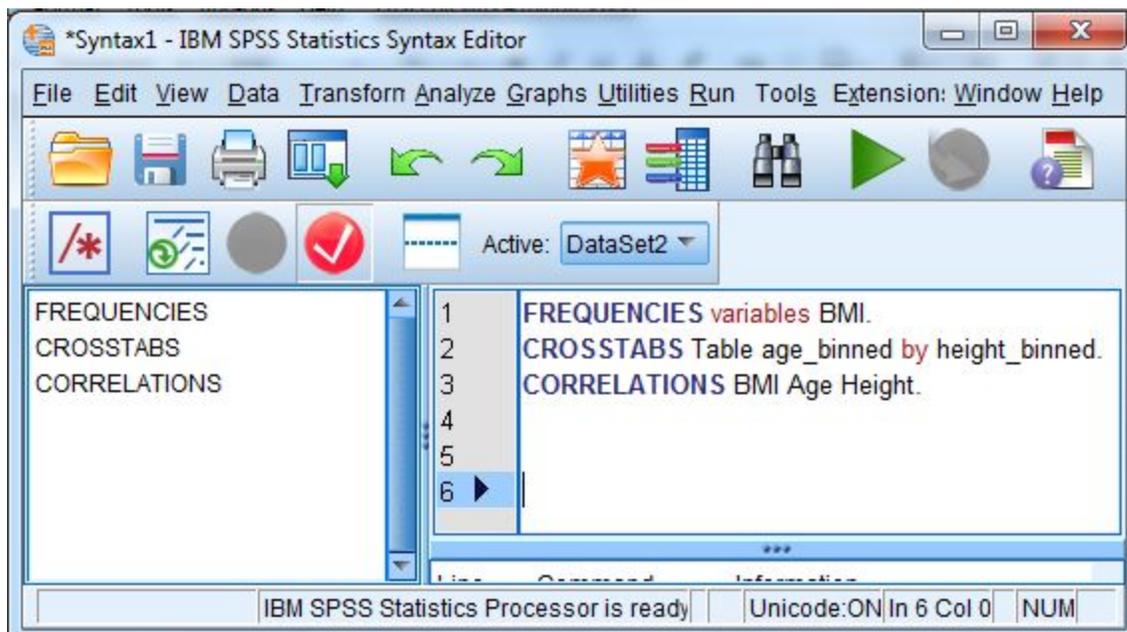
Now we will work with Data editor view.

Let's talk first with Variable views: The Variable View tab displays information about the variables in our data. We can get to the Variable View window in two ways:

- In the Data Editor window, click the Variable View tab at the bottom.

- In the Data Editor window, in the Data View tab, double-click a variable name at the top of the column. This method has the advantage of taking to the specific variable we clicked.

The different tab under variable views are as follows:

1. **Name**: This indicates the name of variables. The following things we have to keep in our mind when typing variable names.
   a. No Spaces
   b. Cannot start with special characters (#, @, &)
   c. Cannot begin with number
   d. Each variable must be unique
   e. 64 characteristics or less

While changing the name of variable the data will not be affected. We can change the name of variable by double clicking in its name.

2. **Type:** There are different types of variable as follows.

Variable Type

- Numeric
- Comma   Characters: 24
- Dot
- Scientific notation
- Date
- Dollar
- Custom currency
- String
- Restricted Numeric (integer with leading zeros)

(i) The Numeric type honors the digit grouping setting, while the Restricted Numeric never uses digit grouping.

OK   Cancel   Help
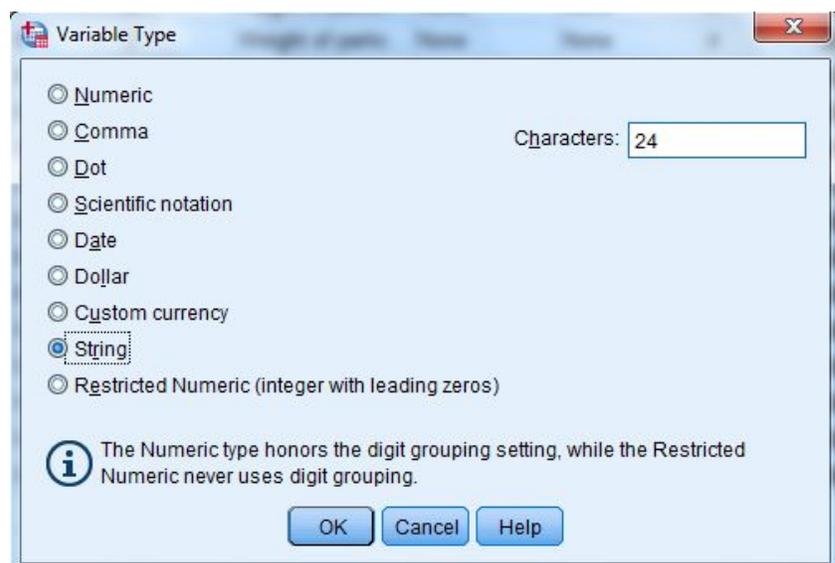
To change a variable's type, click inside the cell corresponding to the "Type" column for that variable. A square "..." button will appear; click on it to open the Variable Type window. Click the option that best matches the type of variable. Click OK.

3. **Width:** It is the number of digits displayed for numerical values or the length of string variable.

4. **Decimals:** It indicates the number of digit to display after a decimal point for values of that variable . This field will not appear for string variable. This changes how the numbers are displayed, but not change the values in dataset.

5. **Label:** A brief but descriptive definition or display name for respective variable.

6. **Values:** For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3). It is strongly suggested that to give each value a label so that you (and anyone looking at your data or results) understands what each value represents.
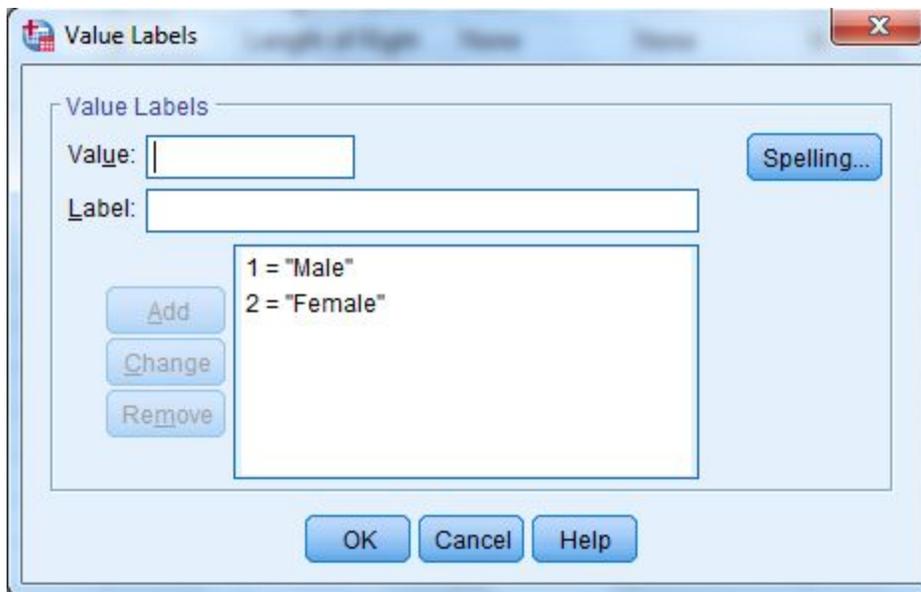
7. **Missing:** Not defining missing value does not affect or eliminate SPSS's default missing value code for numeric variables (".") . However this column allows to specify upto 3 unique missing value code or specific range of data to treat as missing plus one unique missing value code.

8. **Columns:** This indicates the width of column in data view spreadsheet. This is not same as number of digit displayed in each value. This is only used for view purpose.

9. **Align:** This indicates the alignment of content in the cells of the SPSS Data View spreadsheet. Options include left-justified, right-justified, or center-justified.

10. **Measure:** The level of measurement for the variable (e.g., nominal, ordinal, or scale).Some procedures in SPSS treat categorical and scale variables differently. By default, variables with numeric responses are automatically detected as "Scale" variables. If the numeric responses actually represent categories, we must change the specified measurement level to the appropriate setting.

Measurement Level ICON: displayed in dialog box with variable

	Numeric	String	Date	Time
Scale (Continuous)		n/a		
Ordinal				
Nominal				

11. **Role:** The role that a variable will play in your analyses (i.e., independent variable, dependent variable, both independent and

dependent). Some options in SPSS allow you to pre-select variables for particular analyses based on their defined roles. Any variable that meets the role requirements will be available for use in such analyses. You can choose from the following roles for each variable:

Input: The variable will be used as a predictor (independent variable). This is the default assignment for variables.
Target: The variable will be used as an outcome (dependent variable).
Both: The variable will be used as both a predictor and an outcome (independent and dependent variable).
None: The variable has no role assignment.
Partition: The variable will partition the data into separate samples, use for testing and validation.
Split: This function is not used in our SPSS; it is used with the IBM® SPSS® Modeler.

Data Entry in SPSS

After setting all the variables in variables view, we can entry the required data in Data view. When the Data Editor window is accessed for the first time, the top cell of the leftmost column will be highlighted (i.e. thickened black borders round the cell). This is the active cell. We can make any cell active by moving mouse to the required cell and then clicking the left mouse button. We change the active cell, the cell editor on the left, track the location of the active cell. A value typed in from the keyboard will appear in the cell editor and can be transferred to the active cell by pressing return or enter key on the keyboard. We can change position of the active cell in grid by using the cursor keys (i.e. the up, down, right and left arrows on the keyboard). We can now enter data into any cell.

Now we will practice: setup variables and enter data in SPSS.

Exercise 1 – Sample Questionnaire and Coding, Variable Labels, Value Labels and Data entry.

Questionnaire:

1. What is your gender? ……..Male   …….. Female
2. What is your date of birth? DD/MM/YYYY
3. What is your weight in Kg? ………… Kg
4. What is you height in cms? ………..cms
5. Can you tell the number of members in your family? ……..
6. Can you tell me your salary per month?   NPR ………………
7. What is your education?

    ☐ Below SLC   ☐ SLC   ☐ +2 level
    ☐ Bachelors   ☐Masters

Now your turn to enter the variables in variables view and data in data view the following informations.

gender	DOB	weight	height	#family member	salary	Education
Male	15/12/2002	49	150	3	NPR 15,620	Bachelor
Female	18/01/1998	55	145	5	NPR 45,320	Masters
Male	11/12/2003	72	160	8	NPR 55,241	SLC
Male	16/01/1999	65	152	4	NPR 58,621	Below SLC
Male	15/12/2004	49	135	2	NPR 33,521	Masters
Male	9/1/2000	80	125	5	NPR 32,541	Masters
Female	12/12/2005	82	147	8	NPR 11,250	Below SLC
Female	19/01/2001	62	158	6	NPR 65,482	Masters
Female	21/12/2006	58	168	4	NPR 85,498	Masters
Male	25/01/2002	49	147	5	NPR 35,214	Bachelor
Female	15/12/2007	55	155	2	NPR 33,521	Bachelor