Unit 30

Using SPSS for Data Analysis Contents

Contents

Learning Objectives

It is expected that after going through Unit 30, you will be able to

- ❖ Understand the use of SPSS in your data analysis
- ❖ Start and Exit SPSS program
- ❖ Enter the data into a SPSS Data Editor
- ❖ Import a data file from Excel Program

30.1 Introduction

SPSS (Statistical Package for Social Sciences) computer software program provides access to a wide range of data management and statistical analysis procedures. This program can perform a variety of data analysis including tables, statistical analysis and graphical presentation of data. Also, SPSS is particularly well suited to sample survey research.

It is assumed here that you have a basic understanding of the basic concepts and techniques of statistical analysis (see the Units of Block 5 in Book 2). You have also learned how to collect, edit and code data for analysis. In Unit 30, you will learn how to use the SPSS to perform data analysis.

You can run SPSS program on a Personal Computer (PC) within the Windows (95, 98, 2000, XP, or NT) operating system. Since it is a windows based program, you can use the program without any difficulty and more interactively like Word, Excel, or PowerPoint programs. The command instructions given and examples shown in this Unit are Windows based SPSS version 11.5.

Please note that Unit 30 does not carry any information in boxes as it has plenty of graphics to understand the details without the aid of boxes. For Reflection and Action exercises, there are some straight questions for you to answer as the reflection part of the exercise is going to take place during your reading of the text along with its graphics. It is a good idea for you to repeat the viewing of these graphics as many times as

possible and practice using them for carrying out first simpler and later more complex tasks.

30.2. Starting and Exiting SPSS

Normally, SPSS program will be located in the **Programs** folder of your PC. To start the SPSS,

1) Click the left mouse button on the **Start** button located at the lower left of the screen. A number of items will be listed on the screen.

2) Select **Programs**. The program menu will open.

3) Select **SPSS for Windows** from the programs menu and then select **SPSS 11.5 for Windows** from the SPSS menu. Click and release the mouse button. Symbolically, these actions are shown as: select **Start→Programs→SPSS for Windows→SPSS 11.5 for Windows** command from the Start button of your PC. (Throughout this Unit and Unit 32, we will be showing the symbol→to indicate the director (steps)you have to move your cursor with mouse.).

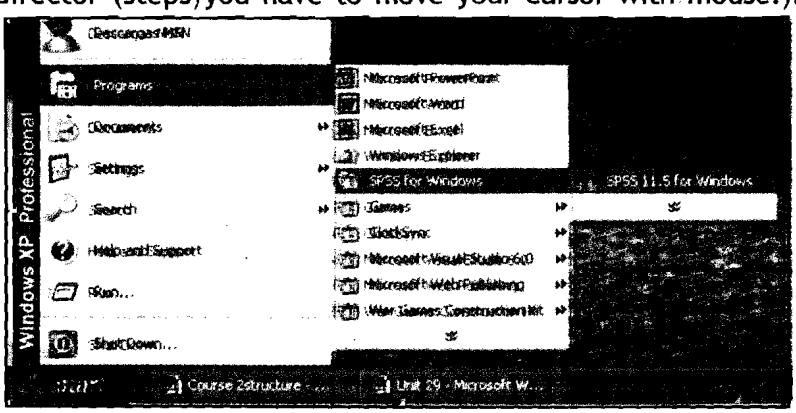


Figure 30.1 SPSS for Windows

4) After a few moments, you will see the **Data Editor** window dialog box along with a **SPSS for Windows** menu dialog box asking you "What would you like to do?".

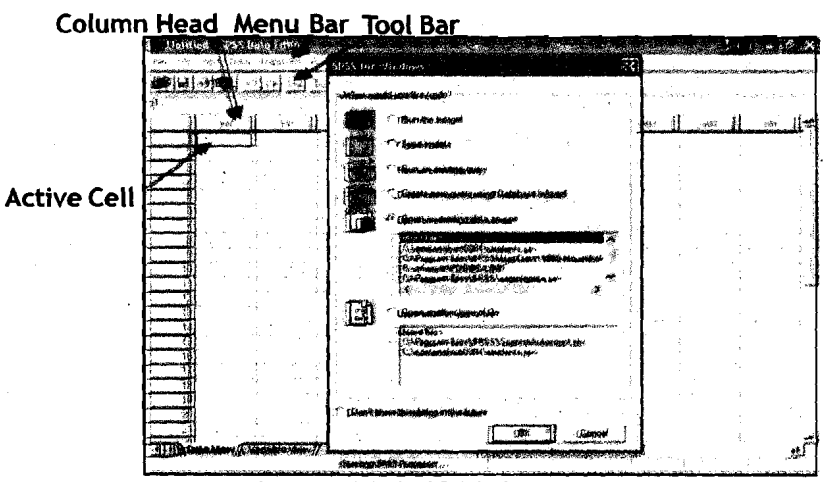


Figure 30.2 SPSS for Data Editor

Exit SPSS Data Editor: Whenever you have finished using SPSS and want to quit it, then select File?Exit command on the menu bar.

30.3 Creating a data file

Normally, the first thing you would like to do is to create a data file. For this, check mark on the [redacted] box and then click OK button on the SPSS for Windows menu dialog box. The menu dialog box disappears from the screen leaving the **Data Editor** on the screen.

Data Editor: The Data Editor helps you:

1) To enter a series of data you have in a specified format required for data analysis.

2) Open an existing file.

3) Edit the data.

4) Converting other data files into SPSS data files.

5) Will be active throughout your session of using SPSS data entry and data analysis.

The Data Editor looks like a worksheet made up of a series of rows and columns. The intersection of a row and column is called a cell. The cells may contain numbers or text. Each column will contain information/ data for each variable. Similarly, each row contains information/data for each case.

The first row of the cells located at the top of each column is shaded and contains a faint **Var**. These cells contain the names of variables. Similarly, the first shaded column contains faint numbers (1,2,3,...). These are called case numbers.

The Data Editor dialog box contains a Menu bar at the top of the window. The menu bar identifies broad categories of SPSS's features called commands. This menu bar helps you in defining and selecting commands.

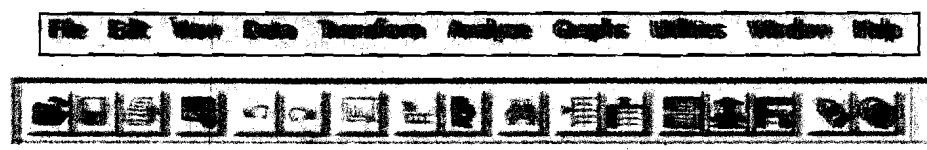


Figure 30.3 Menu bar and Tool Bar of SPSS Data Editor

The Toolbar is below the menu bar and allows you to quickly access basic SPSS commands. By clicking on the respective buttons you can access some commands which will interest you quite often.

Observe the cell at the intersection of row1 (Case 1) and column1 (Var 1) with a heavy border. The heavy border cell indicates that the cell is an active cell. You can enter or edit data in the cell. You can activate any cell in the worksheet by simply pointing the mouse cursor® at it and clicking once.

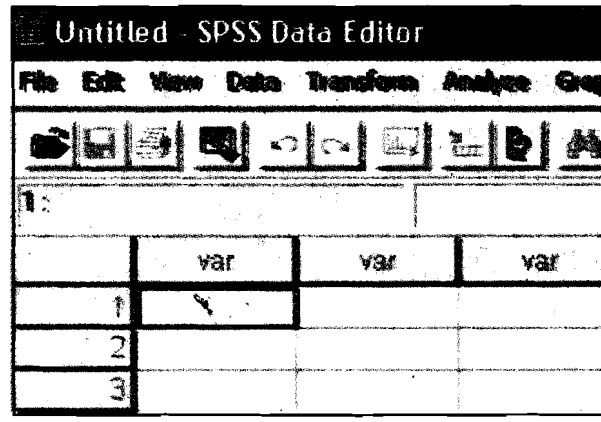


Figure 30.4
Untitiled SPSS Data View

There are two views available in the Data Editor: Data View and Variable View. In the Data View, you can see the data the way you have typed. In the Variable View, you can see the properties of each variable defined. To access these views click on the respective buttons located at the left bottom of the Data Editor screen.

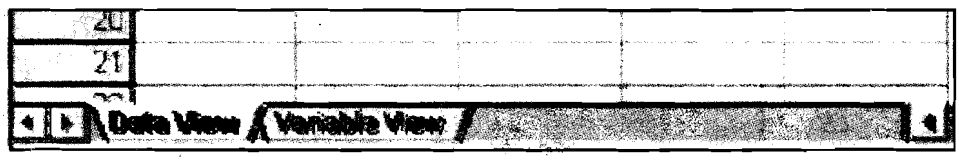


Figure 30.5 Untitiled SPSS Variable View

Entering data into Data Editor

As you have learned earlier, your data may contain a number of variables. Examples of variables are sex (male or female), marital status (married, unmarried, widowed, etc.), income, attitude (likes, dislikes, etc.), a score, etc. Again, for each variable you may have a number of observations called cases. You have also learned in an earlier unit how to assign the codes to the qualitative data. For example, you have a data set containing information of 30 people on sex distribution, age, marital status, and income. You also assigned the following codes for the data set.

Sex distribution: male = 1, female = 2
Age in years
Marital status: married = 1, unmarried = 2, widowed = 3
Income in Rs.
The data you might have collected may look like as follows:

Table 30.1
Data Set by Sex, Age, Marital Status and Income for 30 Persons

Case Number	Sex	Age	Marital Status	Income(Rs.)
1	2	24	1	150000
2	1	52	1	345000
3	2	65	3	45000
4	1	35	3	245000
5	1	42	1	23000
6	1	25	1	670000

7	2	23	2	345000
8	2	63	1	156000
9	1	41	2	65300
10	2	48	1	150000
11	1	34	2	354000
12	2	55	3	23000
13	1	28	1	452000
14	1	43	2	120000
15	1	23	2	456000
16	2	65	1	765000
17	2	67	3	235000
18	2	32	2	54000
19	1	30	2	200000
20	2	25	2	180000
21	1	47	1	210000
22	2	36	3	350000
23	2	70	1	42000
24	1	67	3	175000
25	2	24	2	45000
26	1	32	1	234000
27	1	20	2	125000
28	2	25	1	36000
30	2	40	3	560000
30	2	45	1	234000

The process of data entry into the Data Editor involves four basic steps:

1) Define variables
2) Define labels
3) Define missing values
4) Enter the data into the cells

We will explain these steps with the help of the data given in our earlier example. For this move your cursor to the left bottom of the Data Editor and click on Variable View button, if the Data Editor is not in the Variable View.

Step-1: Define Variables: You need to define a variable to name it, specify the data type (qualitative, quantitative, number of decimal places, etc.), assign labels to the variable and data values, define missing values, and specify levels of measurement (nominal, ordinal, interval/ratio scale). In addition, you can also define the column format. For this,

1. Activate a cell in the first column by clicking on it.
2. Click on the Variable View button. The grid will change to a new format as shown below. For each variable you create, you need to specify all or most of the attributes described by column headings.
3. Activate the first cell in row 1 under **Name** column heading to

change the variable name. Type the name of the variable say Sex and then press enter key. Observe that the variable name Sex replaces with the default variable name Var and in the other cells of the first row the default properties of the variable will appear.

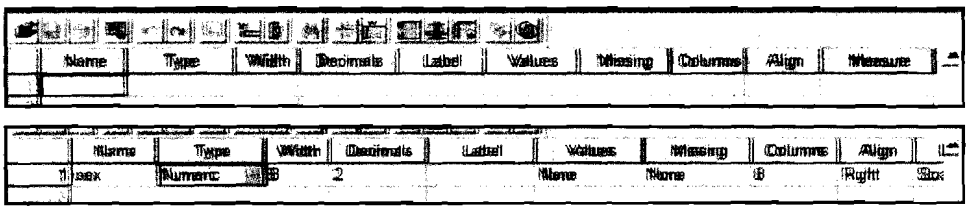


Figure 30.6 Changes in Data Grid and Data Variable

Remember that a variable name can have a maximum of 8 characters containing only letters or both letters and numbers. The first character of a variable name should start with a letter when it contains numbers. There should not be any special characters (like &, ?, !, ',*) in the variable name. Also, in the same data file no two variables should have the common names.

4) To change the type of a variable, move the cursor to the second cell of the first row under **Type** column heading. A small grey button marked with three dots will appear. Click on it. The **Variable Type** dialog box appears on the screen. Notice that Numeric is the default Variable Type. If you have only numeric values for that variable (say Sex variable) check mark the **Numeric** box. You can enter the width of the number (the default width is 8 characters) in the **Width** text box. Sometimes you may need to enter numbers with decimal places. Enter the number of decimal places in the **Decimal Places** text box. The default setting is 2 decimal places. If your data contained only integer values, type 0 in the **Decimal Places** text box. If you have a variable string characters (like names of people, places, etc.), check mark the String button in Variable type dialog box and enter the maximum number of characters that particular string variable can hold. Similarly, other data variable types such as date, currency, etc., can also be defined.

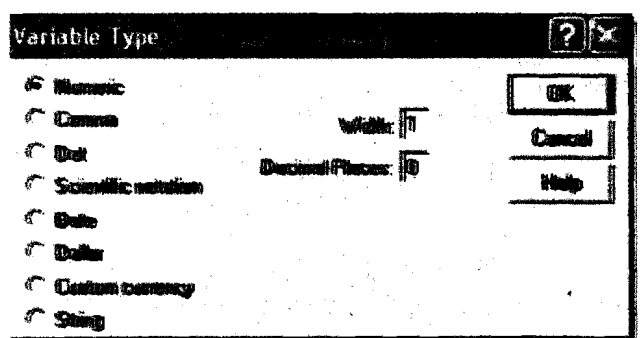


Figure 30.7 Variable Type

5) Specify the level of measurement for the variable (for example, Sex is a nominal variable). By clicking the cell under **Measure** column heading.

Step-2: Define Labels: Now you can assign the text labels to the coded values of the variables. A variable label is a longer description of the

variable that can be included in the variable name you have defined earlier: This may be necessary since the variable name is restricted to only 8 characters and at later stages to understand the characteristics of that variable. To define labels,

1) Type the name of the label (say Sex distribution of persons) in the cell under Label column heading.

2) Move on to the cell under **Values** column heading. Click the grey box with three dots. **Value Labels** dialog box appears on the screen. Type the numerical value assigned for the label under Value text box and type label name for that value under **Value** Label text box. For example, you may type **1** in the Value text box and **Male** in the **Value Label** text box. To add the label, click **Add** button. Again type 2 in the Value text box and Female in the Value Label text box and then click Add button.

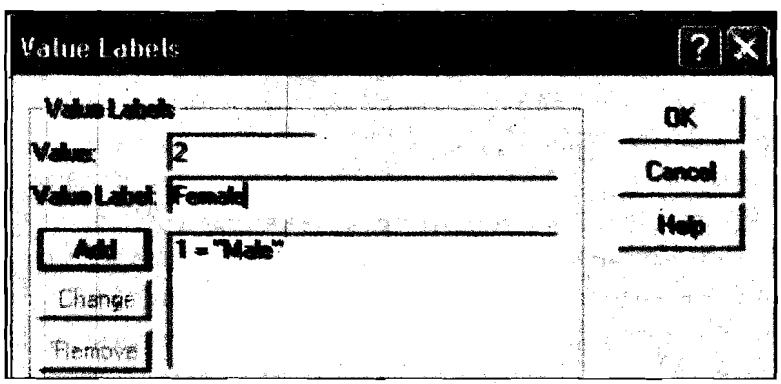


Figure 30.8 Variable Labels

3) Continue this process until you add all the values and the respective labels.

4) Click **OK** button to close the Value Labels dialog box.

Remember that you need to define Value Labels only for categorical data. For the continuous data this is not required.

Step-3: Defining Missing Values: Sometimes, your data may contain missing responses for a variety of reasons. Assign a missing value to the variable if necessary. For example, if the Sex category of a person is not available, you may assign the value 9 to indicate the missing value. The missing value indicates to SPSS that the response is not available and should not be included in the data analysis. To define the missing values to the variable,

1) Click on the cell under **Missing** column heading. The **Missing Values** dialog box appears on the screen.

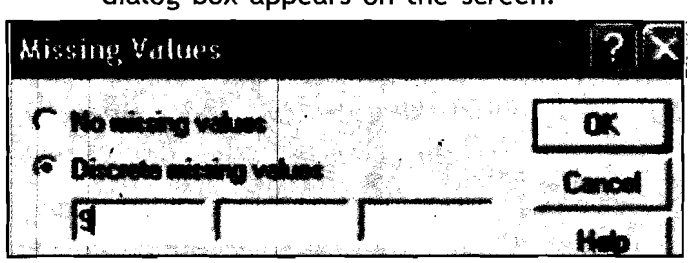


Figure 30.9 Missing Values

2) If there are no missing values in the variable, check mark **No missing values** button. Otherwise check mark **Discrete missing values** button. You can type a number assigned by you as Missing value. For example, you may assign 9 if Sex category of a person is not available. If you have check marked for missing values, type the assigned missing value (say 9) in the text box.

3) Click **OK** button to close the dialog box.

Remember that the missing values you have assigned for a variable are present in their respective positions in a data file.

You can also specify the width of each column in the cell under **Column** column heading and the alignment (Right, Left, and Center) in the cell under **Align** column heading. Specify the type of measure in the cell under Measure heading.

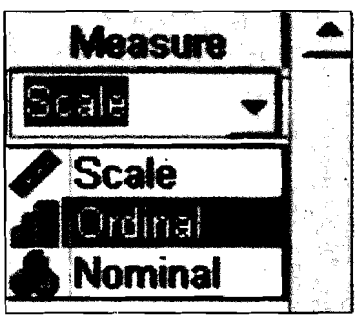


Figure 30.10
Specifying under Measure

Now you have defined all the information for the variable. Move on to the second row (to define 'age' variable), the third row (to define 'marital status' variable), ...etc. to define the rest of the variables in that order.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	sex	Numeric	1	0	Sex distribution	{1, Male}	9	8	Right	Nominal
2	age	Numeric	1	0	Age in years	None	99	8	Right	Scale
3	marital	Numeric	1	0	Marital status	{1, Married}	9	8	Right	Nominal
4	income	Numeric	5	0	Income in Rs.	None	99	8	Right	Scale
5										

Figure 30.11(a) Defining all the variables

you have defined all the variables, you may like to see the generated variable definitions. For this, select **Utilities→File Info** command on the menu bar. This will generate file information in the output window that can be printed if you need for future reference.

File Information

List of variables on the working file

Name
Position

SEX **Sex distribution of persons**

1) Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1
Missing Values: 9

	Value	Label
	1	Male
	2	Female

AGE **Age in years**

2) Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1
Missing Values: *

MARITAL **Marital status**

3) Measurement Level: Nominal
Column Width: 8 Alignment: Right
Print Format: F1
Write Format: F1
Missing Values: 9

Value	Label
1	Married
2	Unmarried
3	Widowed

INCOME **Income in Rs.**

4. Measurement Level: Scale
Column Width: 8 Alignment: Right
Print Format: F6
Write Format: F6
Missing Values: 99

Step-4: Entering the Data into the Cells: Once all the variables for your data file are defined, the data can be directly entered into the cells. For this, first change the view to Data View by clicking the **Data View** button.

1) Click on cell 1 of the (Sex) variable to activate the cell.

2) Type the value of the variable (say '2') and then press **Enter** key.

Observe that now number '2' appears in cell 1 and cell 2 (the cell below cell-1) becomes active.

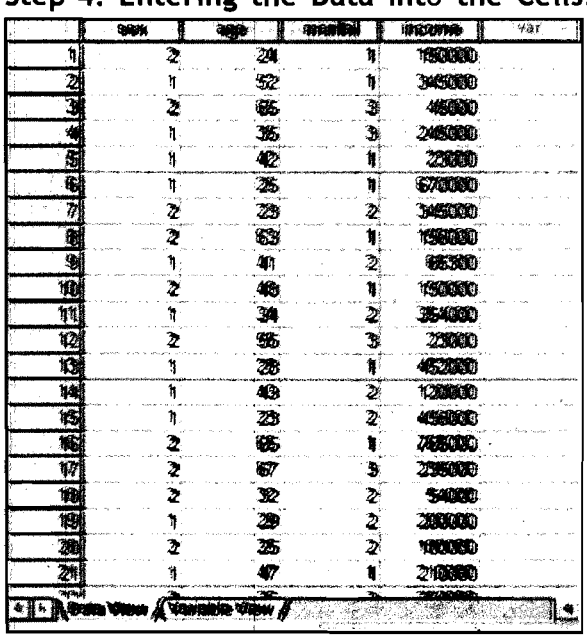


Figure 30.11(b)
Entering Values for Each Variable

3) Type 1 and press Enter key. This indicates that the value for case of the sex variable is also entered. Continue this procedure until all the values of the 30 cases are entered for sex variable.

4) Activate the case 1 cell below Age variable (column) and start entering values for that variable. Continue this procedure until you enter the data for all cases and all variables.ng a data file .

Once you have entered the data, it is a good practice to save the data in a file. This will avoid not only repeating the data entry but also for all future uses of your data. SPSS distinguishes between two types of files: data files (with extension .sav) and output files (with extension .spo). The data files contain the data you have entered. The output files contain the output of the data analysis you have performed. You need to save these files in case you may need them for future use.

To save a data file,

1) Select File→Save As... command from the menu bar. The :Save Data As dialog box appears on the screen.

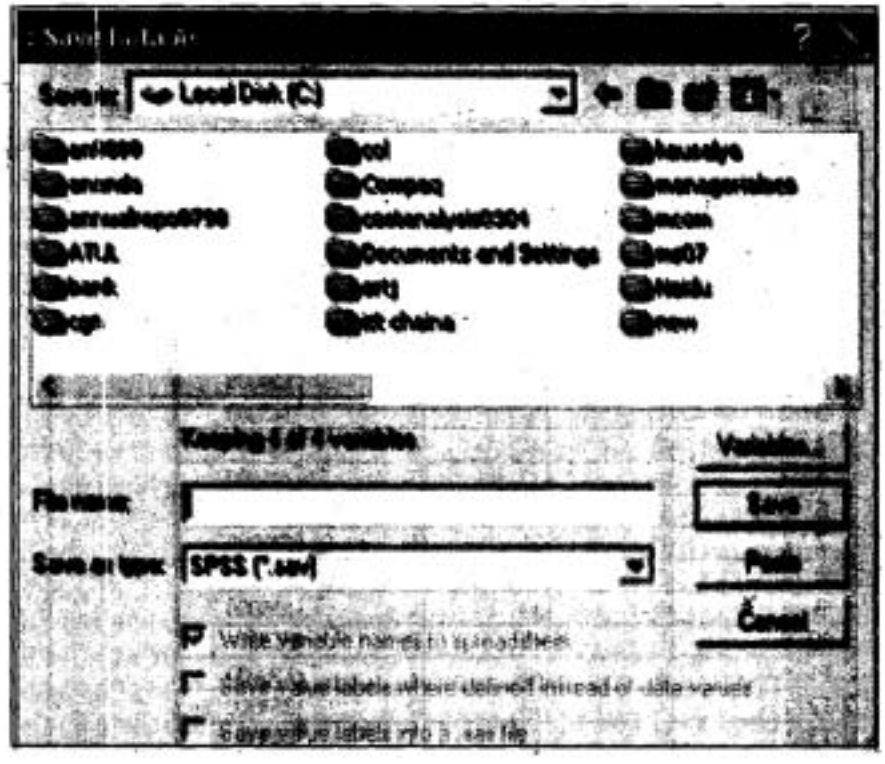


Figure 30.12 Saving a Data File

2) Select the drive and folder where you are interested that your file should be located.

3) Type the file name in the text box under File name box. Click Save button.

Importing a data file from Excel worksheet

Often you might have entered data-in an Excel worksheet and want to use the same data set for analysing the data using SPSS. SPSS can easily

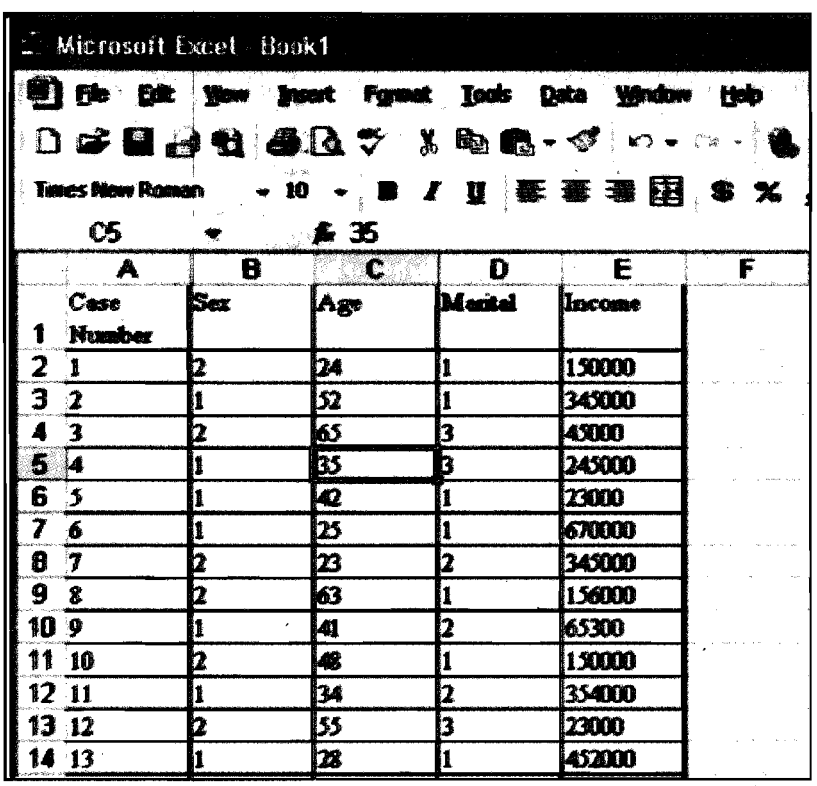
open an Excel data file and some other types of data files. To open an Excel data file,

1) Select **Start→Programs Microsoft Excel** command from the **Start** button on your PC screen. In a few moments the Excel worksheet dialog box appears on the screen.

2) Select **File→Open...** command from the menu bar. The **Open File** dialog box appears on the screen. Select the drive and folder where the data file is stored. Select (or type) the file name and click OK button. The Excel data file will open as shown here.

Observe that the variable names are at the top row. Let us assume that the worksheet has been saved as Excel (with extension .xls) file called Profile.

3) Select **File→Open→Data...** command in the SPSS Data Editor dialog box. The **Open File** dialog box appears. Choose the appropriate directory and folder in **Look in** File of type box. Select the file name Profile.xls in File name box.



	A	B	C	D	E	F
1	Case Number	Sex	Age	Marital	Income	
2	1	2	24	1	150000	
3	2	1	52	1	345000	
4	3	2	65	3	45000	
5	4	1	35	3	245000	
6	5	1	42	1	23000	
7	6	1	25	1	670000	
8	7	2	23	2	345000	
9	8	2	63	1	156000	
10	9	1	41	2	65300	
11	10	2	48	1	150000	
12	11	1	34	2	354000	
13	12	2	55	3	23000	
14	13	1	28	1	452000	

Figure 30.13 Microsoft Excel - Book1

Observe that the variable names are at the top row. Let us assume that the worksheet has been saved as Excel (with extension .xls) file called Profile.

4) Select **File→Open?Data...** command in the SPSS Data Editor dialog box. The **Open File** dialog box appears. Choose the appropriate directory and folder in **Look** in box. Choose the

Excel (*.xls) in the **File of type** box. Select the file name Profile.xls in **Filename** box.

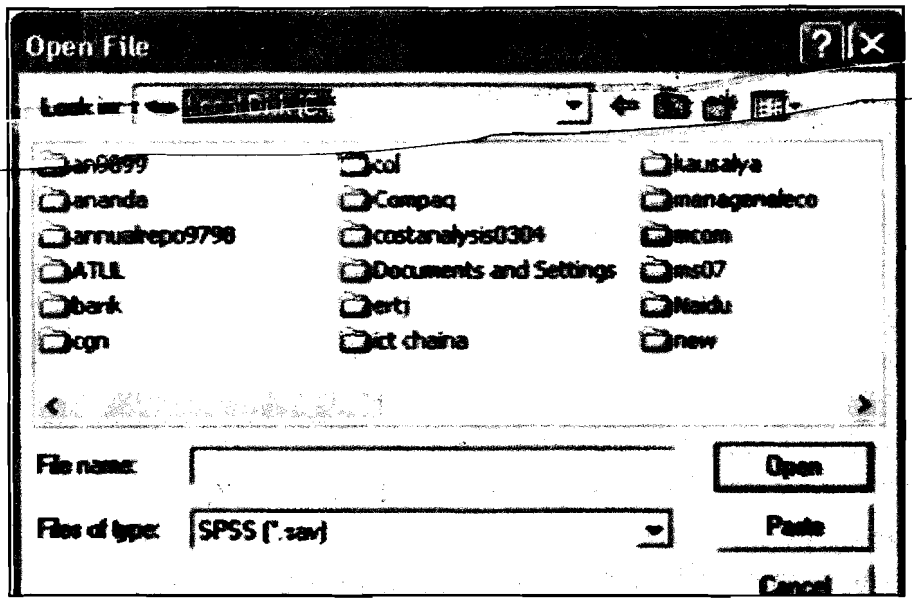


Figure 30.14 Open Files

5) Click **Open** tab in the Open File dialog box. The **Opening Excel Data Source** dialog box appears on the screen. If your Excel file contains variable names, check mark **Read variable names from the first row of the data** box. If you leave the Range box blank, SPSS will read all the available data in the Excel worksheet. If you wish to read only some rows and columns then type a range. For example, you may type **A1:D30** to select first 4 columns (A,B,C, and D) and 30 rows (1 to 30). Click OK button to close the Opening Excel Data Source dialog box and return to the Data Editor dialog box. Observe that the data is in SPSS Data Editor. Save the SPSS data file.

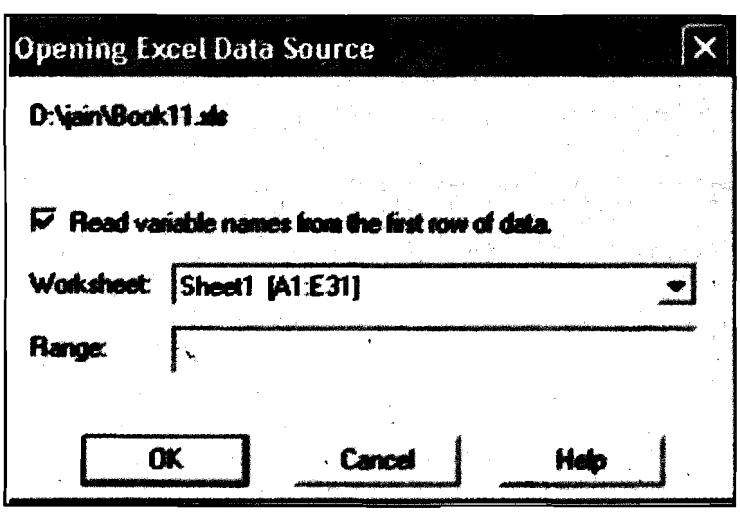


Figure 30.15 Opening Excel Data Source

Reflection and Action 30.1

Answer the following two questions after going through the text above and viewing the graphics. The idea about answering these questions is to make sure that you have understood the procedures.

1. Explain briefly various steps in creating a data file.
2. Differentiate between SPSS Data Editor and Output SPSS Viewer.

30.4 Univariate analysis

Data can be analysed in a number of ways. You can present the data in a simple frequency table or use statistical techniques. For the analysis of data, the data file should be opened first. If the data file is not already open, select the **File→Open** command from the menu bar. The **Open File** dialog box appears. Choose the file location and SPSS data file name and click OK button. The data may now appear in the Data Editor screen.

SPSS offers several tools to analyse the data. The tools are selected from the **Analyze** command on the menu bar. In this section, you will learn how to generate frequency tables and calculate the univariate statistics.

Remember that you can compute the same statistics in SPSS using alternative commands.

Frequency tables

To generate the frequency tables,

1) Select **Analyze→Descriptive Statistics→Frequencies** command from the menu bar. The Frequencies dialog box appears on the screen with two large boxes for variables selection/deselction.

   Observe that the left box contains a list of variables for which data has been entered. On the right side, there is an empty box for the variable(s) that you want to include in the analysis.

2) In the left side box, select the variable to generate a frequency table by clicking the mouse on the name of the variable. Click the arrow ◀ between the two boxes. The selected variable appears on the right side box. Follow these steps till all the variables you want to include for the data analysis appear on the right side box.

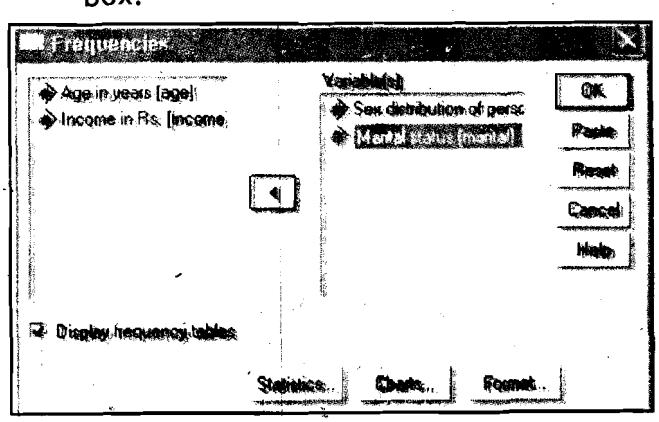


Figure 30.16 Frequencies

3) When the first variable is moved to the right side box an arrow in the opposite direction appears between the two boxes. If you commit a mistake by selecting a wrong variable, click this arrow to return the variable to the original list.

Remember that this Frequencies tool is appropriate only for the categorical data (like Sex and Marital status in our example). Therefore, do not select any continuous variables (like in Age and Income in our example).

4) When you have selected all the variables, you want to include for data analysis, click the **OK** button. In the **Output SPSS Viewer window**, you should now see the output shown below.

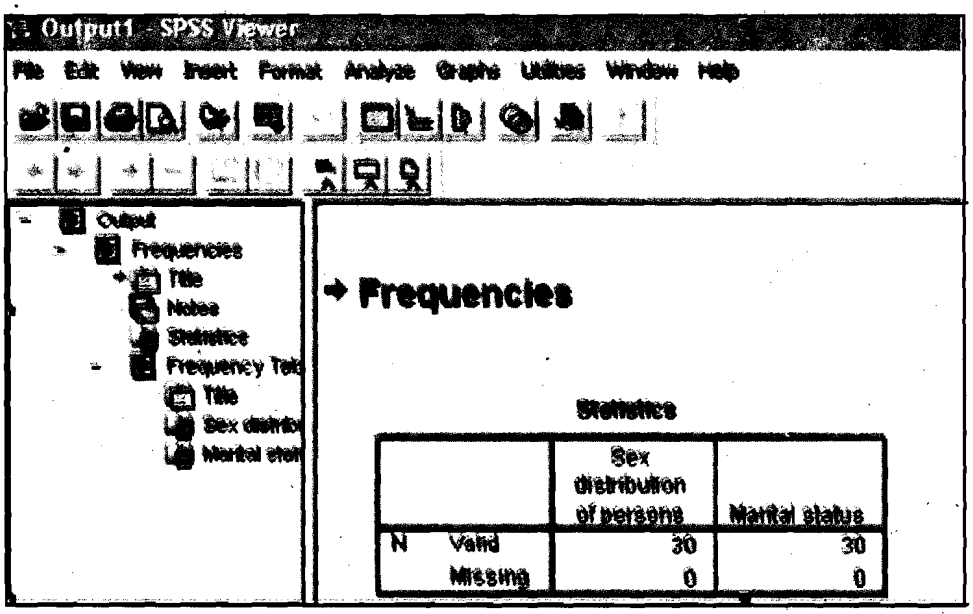


Figure 30.17 Output SPSS Viewer

Frequencies

Find below two tables (Table 30.4 and 30.5 on Sex \Distribution of Persons and Marital Status, respectively).

Table 30.2 Sex\Distribution of Persons

Statistics

		Sex Distribution of Persons	Marital Status
N	Valid	30	30
	Missing	0	0

Table 30.3 Marital Status

Frequencies table

Sex distribution of persons

		Fequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	14	46.7	46.7	46.7
	Female	16	53.3	53.3	100.0
	Total	30	100.0	100.0	

Table 30.4 Marital Status

Marital status

		Fequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	13	43.3	43.3	43.3
	Unmarried	10	33.3	33.3	76.7
	Widowed	7	23.3	23.3	100.0
	Total	30	100.0	100.0	

Observe that the tables contain all the information you want. Also, you have information on the missing cases for each variable. You also have three types of percentages: one for all the cases including missing cases under the heading **Percent**, the second one for only valid cases under the heading **Valid Percent**, and the third gives cumulative percentage under the heading **Cumulative Percent**.

Printing/Saving Output: Now you may want to print on paper or save in a file all or part of the output available in the Output SPSS Viewer window.

1) To save the output in a file, follow the instructions given at saving a data file.

2) To print all the output click anywhere in the Outline pane located on the left side of the screen. Select **File→Print** command from the menu bar. Click OK button.

3) To print a part of output, move the cursor to the end of the portion you want to print. Press the **Shift** key on the keybord and click the left mouse button. Observe that the selection is highlighted. Select File→Print command from the menu bar. Click OK button.

Recode Data: you may want to recode your data for a variety of reasons. For example, the data values for the variable Age are continuous. Now you may want to group them like,

Old Value	New Value
Less than 19	1
20-30	2
30-39	3
40-49	4
50-59	5
60 and above	6

Table 30.5 Data values

To recode a variable, select **Tranform→Recode→Into Diffrent Variables** command from the menu bar. The **Recode into Different Variables** dialog box appears on the screen.

Select the variable you want to recode into different variables in the left side box. Transfer this variable to the right side box using arrow tab that lies between the two boxes. Type the name of the variable in the **Name** text box and label name in the **Label** text box.

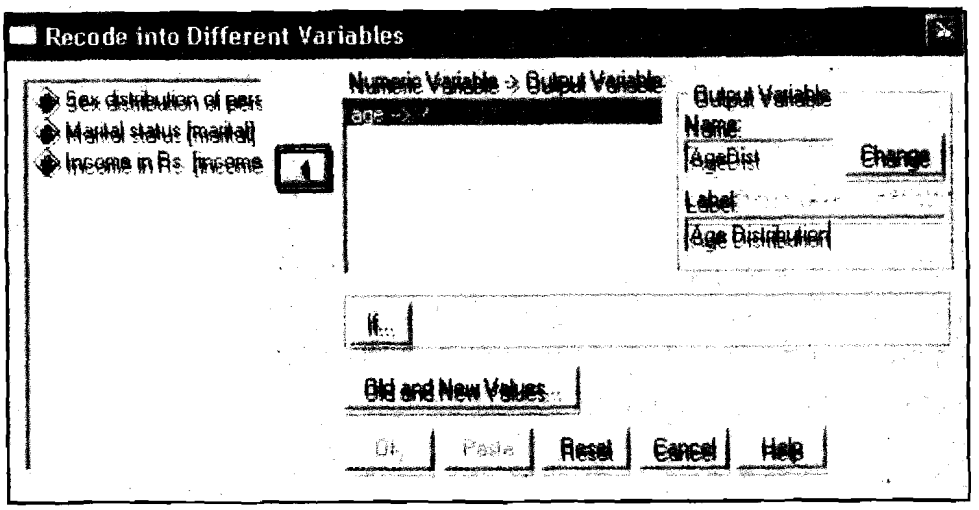


Figure 30.18 Recode into Different Variables: Old and New Values

Press the **Old and New Values** button in the Recode into Different Variables dialog box. The **Recode into Different Variables: Old and New Values** dialog box appears.

Check mark the **Range** button. Type the first range of values in the boxes under Range. Check mark the **Value** button under **New Name** heading. Press Add button to define Old and New values. Press **Continue** button to close the Recode into Different Variables: Old and New Values dialog box and return to the Recode into Different Variables dialog box.

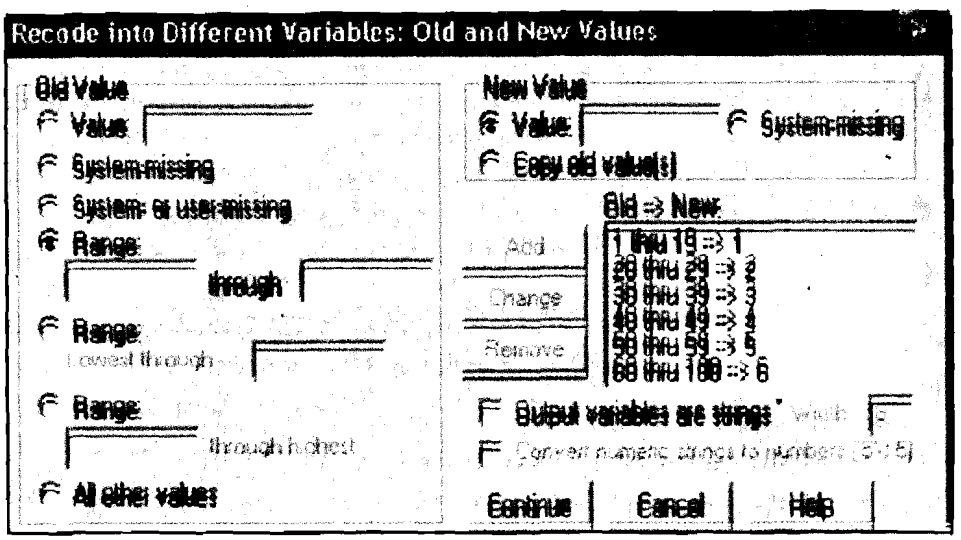


Figure 30.19 Recode into Different Variables

If you have more variable(s) to redefine, continue these steps for each variable. Otherwise press **OK** button to close the Recode into Different Variables dialog box.

Univariate© statistics

For each variable of your data set, you can calculate:

a) Measures of central tendency: Mean, Median, and Mode.

b) Dispersion: Standard deviation, Variance, Range, Standard Error of Mean, etc.

c) Distribution: Kurtosis and Skewness

Remember in SPSS there are some restrictions on the choice of measures of central tendency (Mean, Median, and Mode) that can be calculated on any data set. The choice of Mean, Median, and/or Mode is restricted by the level of measurement of a variable you have defined. If the level of measurement for a variable is nominal, you can calculate only mode. If the level of measurement of a variable is ordinal then you can calculate Mode and/or Median. If the level of measurement of a variable is interval/ratio, you can calculate Mode, Median, and/or Mode.

To calculate the univariate statistics,

1) Select **Analyse→Descriptive Statistics→Frequencies** from the menu bar. The Frequencies dialog box appears on the screen.

2) Transfer the variables on which you want to perform the data analysis from left side box to right side box (as you have done for frequencies analysis earlier).

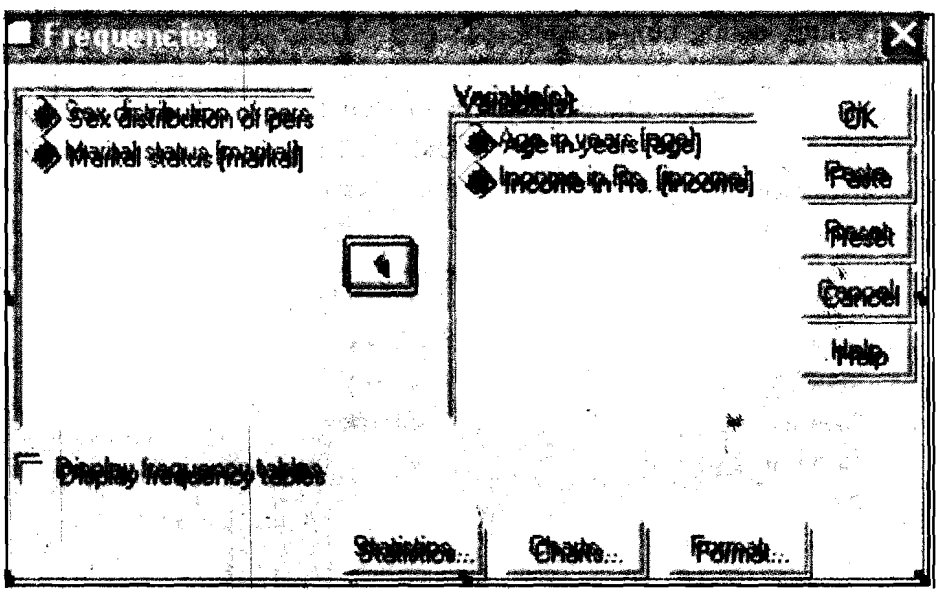


Figure 30.20 Frequencies Analysis

3) If you don't want to display frequencies, remove the check mark in **Display frequency tables** button by clicking. The **SPSS for Windows** dialog box appears asking you to confirm. Click **OK** button to close that window.

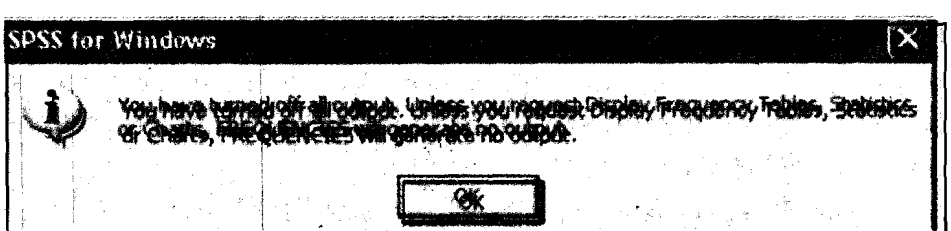


Figure 30.21 Closing SPSS for Windows

4) Click on the **Statistics...** button in the Frequencies dialog box. The **Frequencies:Statistics** dialog box appears on the screen.

5) In the area under **Central Tendency**, check mark the appropriate button to calculate Mean, Median, and/or Mode.

6) In the area under **Dispersion**, check mark the appropriate buttons to calculate Standard deviation, Variance, Range, Standard error or Mean, etc.

7) In the area under **Distribution**, check mark the appropriate buttons to calculate Skewness and/or Kurtosis.

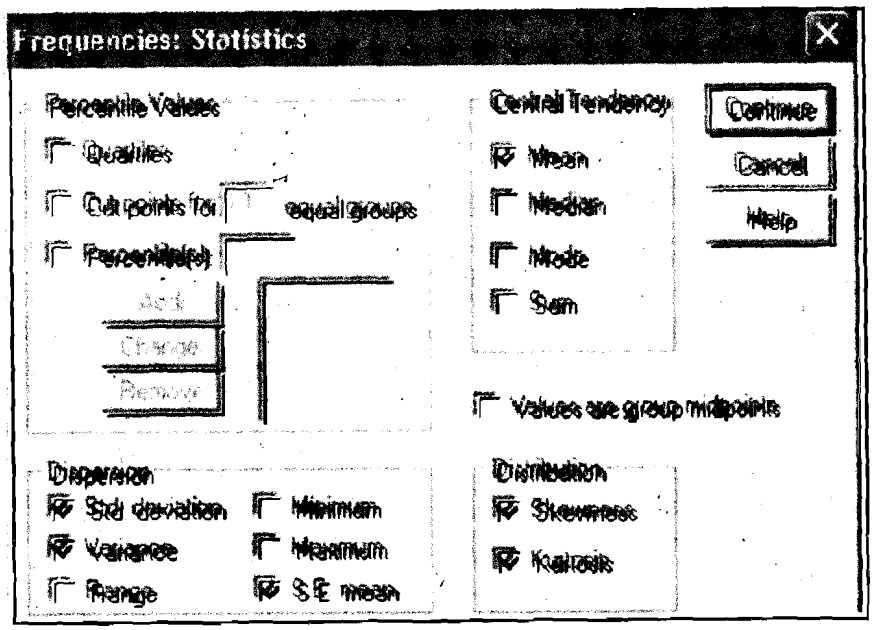


Figure 30.22 Frequencies Statistics

8) Click on **Continue** button to close the **Frequencies: Statistics** dialog box.

9) Click on OK button in the Frequencies dialog box to close it.

Observe that now the output is Descriptive Statistics.

Table 30.6 Statistics

		Age in Years	Income in Rs.
N	Valid	30	30
	Missing	0	0
Mean		40.83	234810.00
Std. Error of Mean		2.891	35206.370
Std. Deviation		15.836	192833.232
Variance		250.764	37184655414
Skewness		.531	1.182
Std. Error of Skewness		.427	.427
Kurtosis		−1.012	1.059
Std. Error of Kurtosis		.833	.833

Remember that you should opt for only appropriate statistics. For example, there is no meaning in opting for Mean of a sex Variable since there is nothing like mean of a sex distribution.

Reflection and Action 30.2

You have just finished reading about univariate analysis in which you worked on frequency tables and univariate statistics. In the light of this information answer the following questions.

- ❖ What is the command on the SPSS menu bar to perform frequencies data analysis?
- ❖ Once the data has been coded and entered into the SPSS Data Editor, is it possible to recode the data? If yes, what is the command to recode?
- ❖ You have defined the level of measurement of a variable as ordinal. Is it possible to calculate all the measures of central tendency for this variable using SPSS. Name the central tendency measures you can calculate.

30.5 Bivariate© Analysis

Often, you may be interested in comparing two sets of data or Variables to explore the relationship between two Variables. In this Section, you will learn how to cross-tabulate the data, coefficient of correlation, linear regression, and coefficient of variance to compare two variables.

Cross-tabulation of data

To capture any possible relationship between two variables measured with categorical data, you may use bivariate table, which is also known as cross-tabulation. To cross-tabulate two variables follow the steps given below.

1) Select **Analyse?Descriptive Statistics?Crosstabs...** command from the menu bar. The **cross tabs** dialog box appears on the screen.

2) Click on the Variable in the source list (from left side box) that will form the rows of the table. Shift this Variable to the box under **Row(s)** using arrow key.

3) Similarly, shift the variable that will form the columns of the table from source list to the box under column(s).

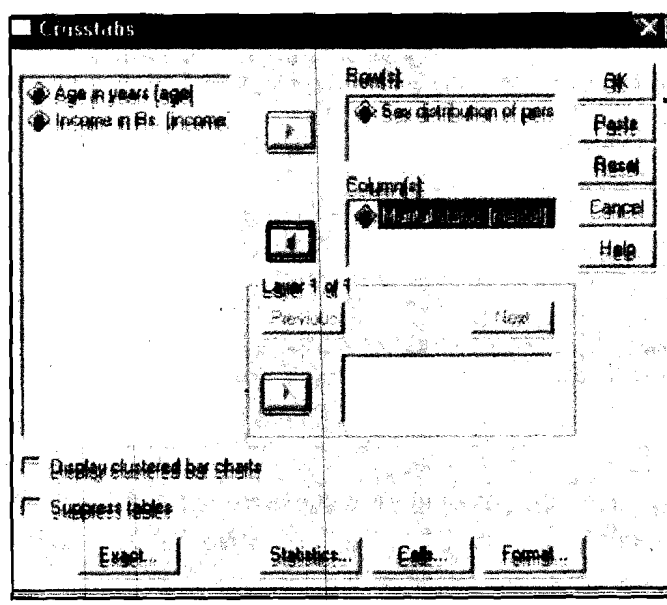


Figure 30.23 Crosstabs

4) Click **OK** button. You will find the table in the Output viewers window.

Crosstabs

Table 30.7 Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex distribution of persons@ Marital status	30	100.0%	0	.0%	30	100.0%

Count

Table 30.8 Sex distribution of persons' Marital status Crosstabulation

		Marital Status			Total
		Married	Unmarried	Widowed	
Sex distribution of persons	Male	6	6	2	14
	Female	7	4	5	16
Total		13	10	7	30

Sometimes, you may be interested in calculating the row/column/total percentages in a table. For this, click on **Cells...**tab in the cross tabs dialog box before step-4 above. The **Crosstabs: Cell Displays** dialog box appears on the screen.

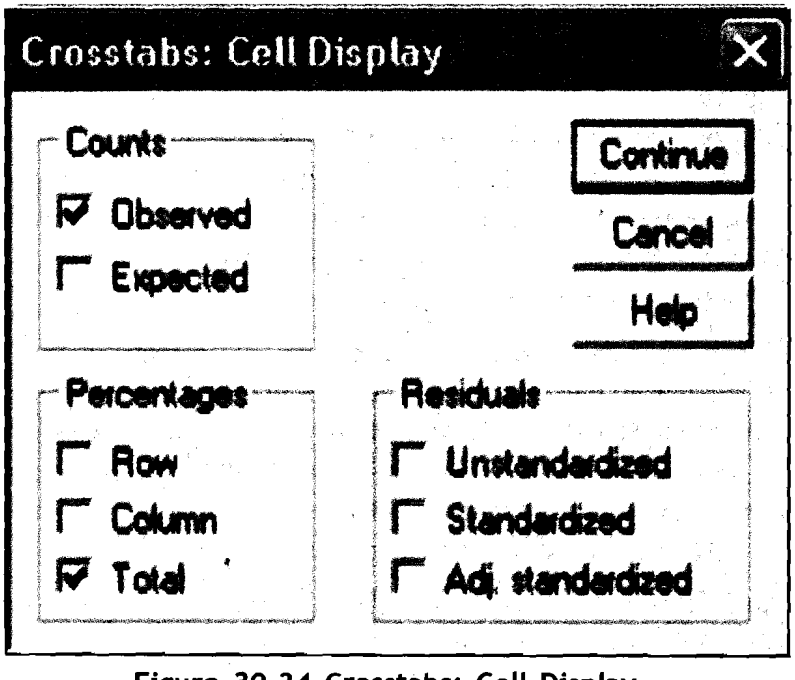


Figure 30.24 Crosstabs: Cell Display

Check mark the appropriate (Row, Column, and/or Total) button under **percentages** area. Click **continue** button to close the Crosstabs: Cell

Display dialog box. Click **Ok** button on the Crosstabs dialog box, to view the output (see Table 30.9 and Table 30.10).

Crosstabs

Table 30.9 Case Processing Summary: Output

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex distribution of persons* Marital status	30	100.0%	0	.0%	30	100.0%

Table 30.10 Sex distribution of persons *Marital status Crosstabulation: Output

			Marital status			
			Married	Unmarried	Widowed	Total
Sex distribution of persons	Male	Count	6	6	2	14
		% of Total	20.0%	20.0%	6.7%	46.7%
	Female	Count	7	4	5	16
		% of Total	23.3%	13.3%	16.7%	53.3%
		Count	13	10	7	30
		% of Total	43.3%	33.3%	23.3%	100.0%

Bivariate statistics

The statistical tools you often use to compare two Variables may be the coefficient of variance, correlation, and linear regression.

Coefficient of variance: As you are aware, the Coefficient of Variance (CV) is the standard deviation expressed as a percentage of the mean.

$$CV = \frac{\text{Standard deviation}}{\text{Mean}} \times 100.$$

Unfortunately, SPSS does not have a command to complete the Coefficient Variance for a variable in a data file. What we advise you is that you should calculate the respective Mean and Standard deviation of a Variable using the Descriptive dialog box as explained earlier and then calculate the CV by hand which is very simple.

Correlation coefficient: There are two types of correlation coefficients: Pearson's correlation coefficient and Spearman's rank correlation coefficient. The Pearson correlation is appropriate and applicable when you have interval/ratio data. The Spearman rank correlation coefficient is applicable when you have two ordinal scales with a large number of values or one ordinal and the other interval/ratio scale.

To compute the appropriate correlation coefficient for your data set, follow the instructions given below.

1) Select **Analyse→Correlate→Bivariate...** command from the menu bar. The Bivariate Correlations dialog box appears on the screen.

2) Select the variables by shifting from left side box to the box under **Variables** area.

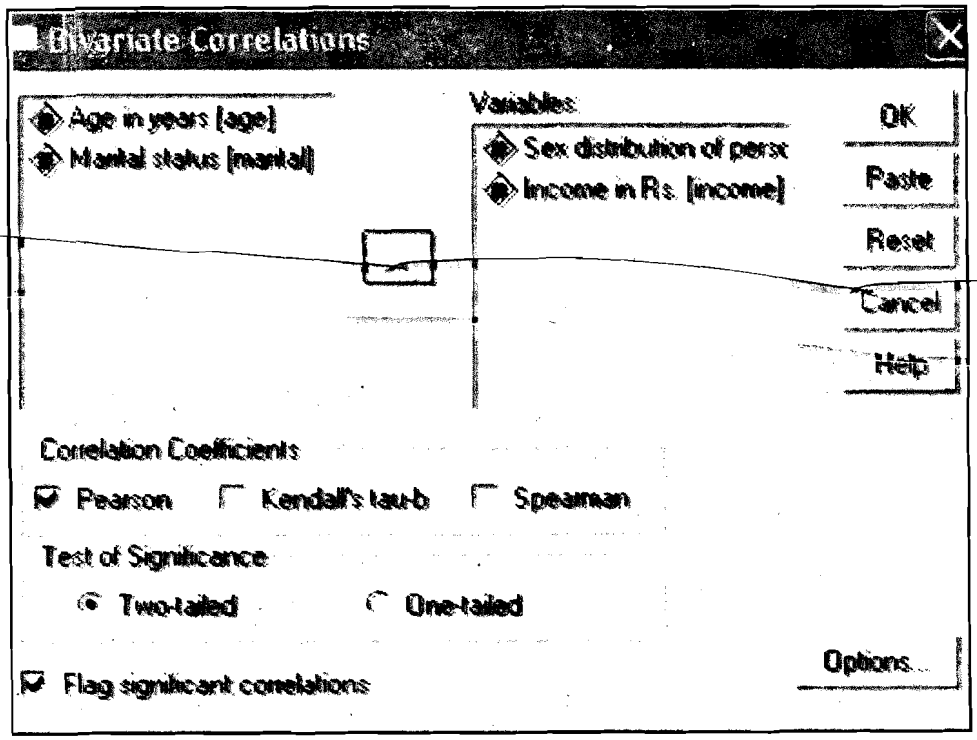


Figure 30.25 Bivariate Correlation

3) Check mark the appropriate (Pearson, and/or Spearman) button under **Correlation Coefficients** area to select the type of correlation coefficient.

4) If you want the associated tests of significance, check mark the appropriate (Two-tailed or one-tailed) button under Test of Significance area. You will see the results in the Output Viewer Window.

Correlations

Table 30.11 Correlations between Two Variables

		Sex distribution of persons	Inncome in Rs.
Sex distribution of persons	Person Correlation	1	−.136
	Sig. (2-tailed)	.	.472
	N	30	30
Inncome in Rs.	Person Correlation	−.136	1
	Sig. (2-tailed)	.472	.
	N	30	30

Linear regression: The linear regression technique is used to: (a) test the hypotheses concerning the linear relationship between two variables (b) estimating the specific nature of relationship; and (c) to predict the values of dependent variable when you know the values of independent variable. To run the linear regression procedure follow the steps given below.

1) Select **Analyse→Regression→Linear**...command from the menu

bar. The Linear Regression dialog box appears on the screen.

2) Click on the variable name that will be **dependent variable** in the left side box. Shift the dependent variable to the box under **Dependent** area using arrow tab.

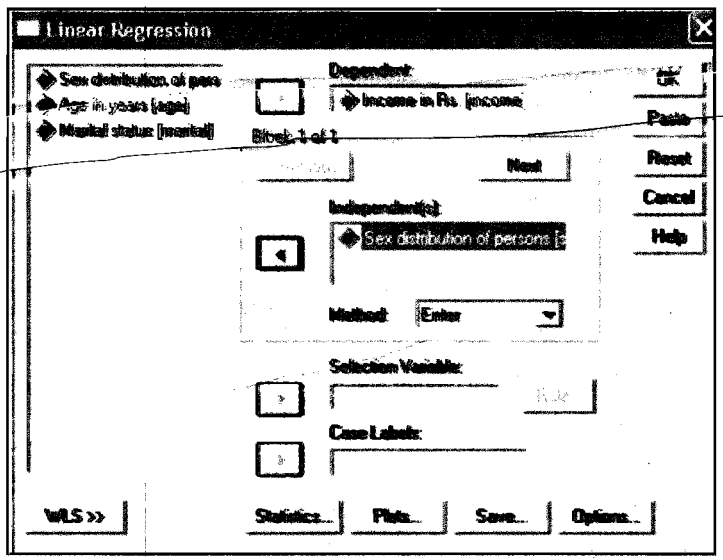


Figure 30.26 Linear Regression

3) Click on the variable name that will be independent in the left side box. Shift the independent variable to the box under **InDependent(s)** area using arrow tab.

4) Click **OK** button. You will see the results in the Output Viewer Window.

Observe that the output consists of four points: (a) a table of variables used in regression, (b) a model summary, (c) an ANOVA table; and (d) a table of coefficients. You may be interested in a portion of the output. We will explain how to select a partial output in another Unit on use of SPSS in report writing.

Regression

Table 30.12 Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Sex distribution of persons	.	Enter

a) All requested variables entered
b) Dependent Variable : Income in Rs.

Table 30.13 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.136[a]	.019	−.016	194413.110

a) Predictors: (Constant), Sex distribution of persons

Table 30.14 ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.01E+10	1	2.005E+10	.531	.472[a]
	Residual	1.06E+12	28	3.780E+10		
	Total	1.08E+12	29			

a) Predictors: (Constant), Sex distribution of persons
b) Dependent Variable: Income in Rs.

Table 30.15 Coefficients[a]

Model		Unstandardzed Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	314275.0	114722.6	−.136	2.739	.011
	Sex distribution of persons	−51825.0	71147.913		−.728	.472

a) Dependent Variable: Income in Rs.

30.6 Multivariate Analysis

Sometimes, you may be interested in exploring complex relationships involving more than two variables. In this section, you will learn how to cross tabulate your data for the analysis of relationships between more than two variables. Similarly, how to compute multiple regression analysis has also been covered.

Elaboration of cross-tables

In the earlier section, you have learned the cross-tabulation of two variables. You can introduce a third variable by sub grouping one of the two variables. This can be done by introducing a variable as control variable. A control variable decomposes the data into sub-groups based on the categories of the control variable. To add a control variable for your cross-tabulation, follow the steps given below.

1) Select **Analyse→Descriptive Statistics→Crosstabs**...command from the menu bar. The **Crosstabs** dialog box appears on the screen as shown on the next page.

2) Click on the variable in the source list that will form the row(s) of the table. Shift this variable to the box under Row(s) using arrow key.

3) Similarly, shift the variable that will form the columns of the table from source list to the box under **Column(s)** using arrow key.

4) Click on the variable which will act as control variable l (this variable splits the variable selected at step-3 into sub groups). Shift the control variable to the box under **Layer 1 of 1** using arrow key.

5) For computing the row/column/total percentages in the table,

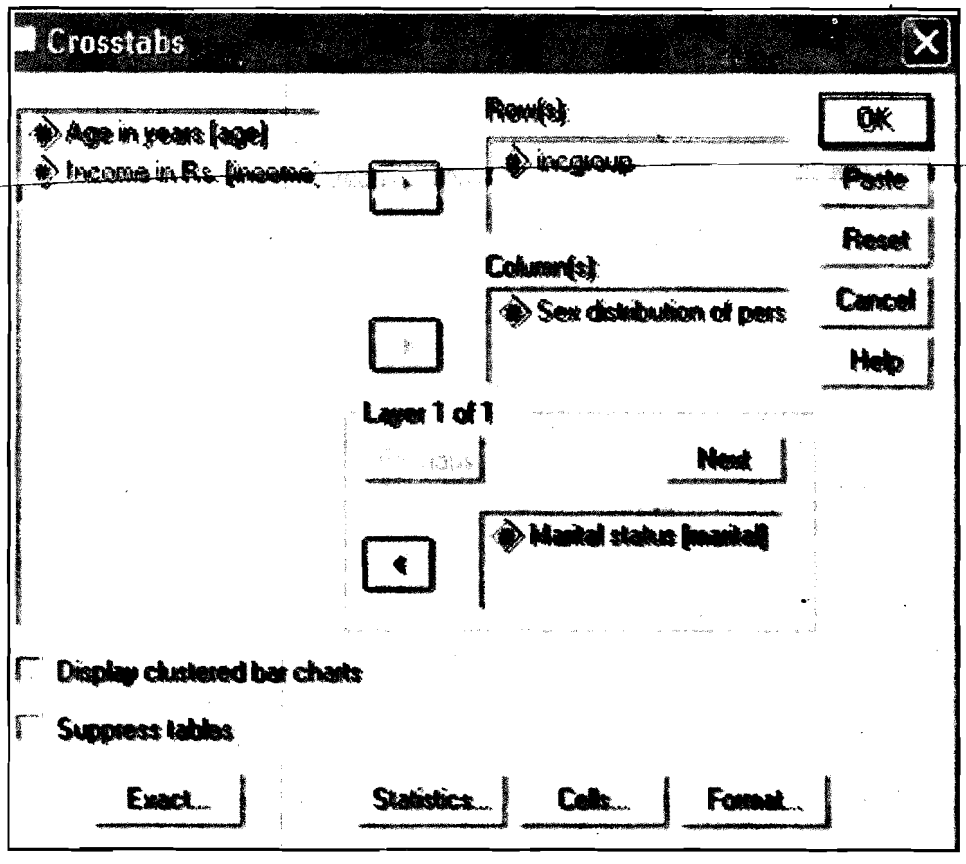


Figure 30.27
Crosstabs: Cell Display for Computing the Row/Column/Percentages

click on **Cells...**tab. The **Cross tabs: Cell Display** dialog box appears on the screen. Click the appropriate (Row, Column, and/or Total) button under **Percentages** area. Click **Continue** button to close the Cross tabs: Cell Display dialog box.

6) Click **OK** button to close the Crosstabs dialog box.

Crosstabs

Table 30.16 Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Income group * Sex distribution of persons * Marital status	30	100.0%	0	.0%	30	100.0%

Let's look at the table 30.17: Income Group and Sex Distribution of Persons' Marital Status Crosstabulation. You will find crosstabulation of marital status by income group of married, unmarried and widowed persons with their sex distribution. This is a good example that you can try and use in your own research work. Not only this but you can also use the other examples given in this unit for your research project.

Table 30.17 Income Group and Sex Distribution of Persons' Marital Status Crosstabulation

Marital status				Sex distribution of persons: Male	Female	Total
Married	Income group	Less than Rs. 50000	Count	1	2	3
			% of Total	7.7%	15.4%	23.1%
		Rs. 100001-200000	Count	0	3	3
			% of Total	.0%	23.1%	23.1%
		Rs. 200001-400000	Count	3	1	4
			% of Total	23.1%	7.7%	30.8%
		More than Rs. 400000	Count	2	1	3
			% of Total	15.4%	7.7%	23.1%
	Total		Count	6	7	13
			% of Total	46.2%	53.8%	100.0%
Unmarried	Income group	Less than Rs. 50000	Count	0	1	1
			% of Total	.0%	10.0%	10.0%
		Rs. 50001100000	Count	1	1	2
			% of Total	10.0%	10.0%	20.0%
		Rs. 100001-200000	Count	3	1	4
			% of Total	30.0%	10.0%	40.0%
		Rs. 200001-400000	Count	1	1	2
			% of Total	10.0%	10.0%	20.0%
		More than Rs. 400000	Count	1	0	1
			% of Total	10.0%	.0%	10.0%
	Total		Count	6	4	10
			% of Total	60.0%	40.0%	100.0%
Widowed	Income group	Less than Rs. 50000	Count	0	2	2
			% of Total	.0%	28.6%	28.6%
		Rs. 100001-200000	Count	1	0	1
			% of Total	14.3%	.0%	14.3%
		Rs. 200001-400000	Count	1	2	3
			% of Total	14.3%	28.6%	42.9%
		More than Rs. 400000	Count	0	1	1
			% of Total	.0%	14.3%	14.3%
	Total		Count	2	5	7
			% of Total	28.6%	71.4%	100.3%

Multiple regression

In two variable linear regression you have used one dependent variable and one independent variable. The multivariate regression is used to investigate the relationship between two or more independent variables on a single dependent variable. The procedure for computing the statistics for multiple regression is the same as that for two variable linear regression explained earlier, except that you have more than one variable under **Independent(s)** area in the **Linear Regression** dialog box.

1) **Analyse→ Regression-Linear**...command from the menu bar. The **Linear Regression** dialog box appears on the screen.

2) Click on the variable name that will be a dependent variable in the left side box. Shift this variable to the box under **Dependent** area using arrow tab.

3) Click on the variable name that will be independent in the left side box. Shift this variable to the box under **Independent(s)**

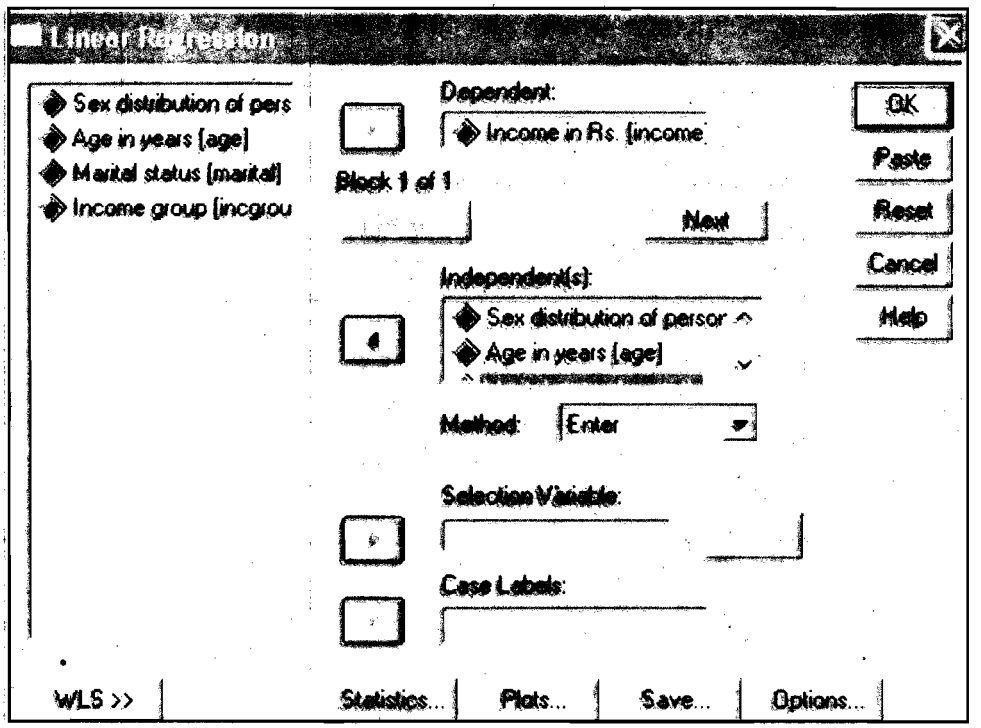


Figure 30.28 Linear Regression Dialog Box

area. Follow this step until all the desired independent variables are selected.

4) Click **OK** button. You will see the results in the Output Viewer Window.

Regression

Table 30.18 Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Marital status, Sex distribution of persons, Age in yeas		Enter

a) All requested variables entered
b) Dependent Variable: Income in Rs.

Table 30.19 Model Summary

Model	R	R Square	Adjusted R Sqaure	Std. Error of the Estimate
1	.168[a]	.028	−.084	200774.054

a) Predictors: (Constant), Marital status, Sex distribution of persons, Age in years

Table 30.20 ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.03E+10	3	1.010E+10	.250	.860[a]
	Residual	1.05E+12	26	4.031E+10		
	Total	1.08E+12	29			

a) Predictors: (Constant), Marital status, Sex distribution of persons, Age in years
b) Dependent Variable: Income in Rs.

Table 30.21 Coefficients[a]

Model		Unstandardzed Coefficients		Standardzed Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	362849.3	153394.2		2.365	.026
	Sex distribution of persons	−44008.8	75727.227	−.116	−.581	.566
	Age in years	−680.236	2431.316	−.056	−.280	.782
	Marital status	−18212.7	46773.389	−.076	−.389	.700

a) Dependent Variable: Income in Rs.

30.7 Tests of Significance

In this section, you will learn one sample t-test* for a mean, two sample t-test for the equality of means, and chi-square test for independence.

One sample t-test

The one sample t-test compares the sample means to a population mean, using t-distribution as the standard of comparison.

1) Select Analyse→Compare Means→One-Sample TTest...command from the menu bar. The **One Sample T Test** dialog box appears on the screen.

2) Click on the variable name you want to perform the t-test on the

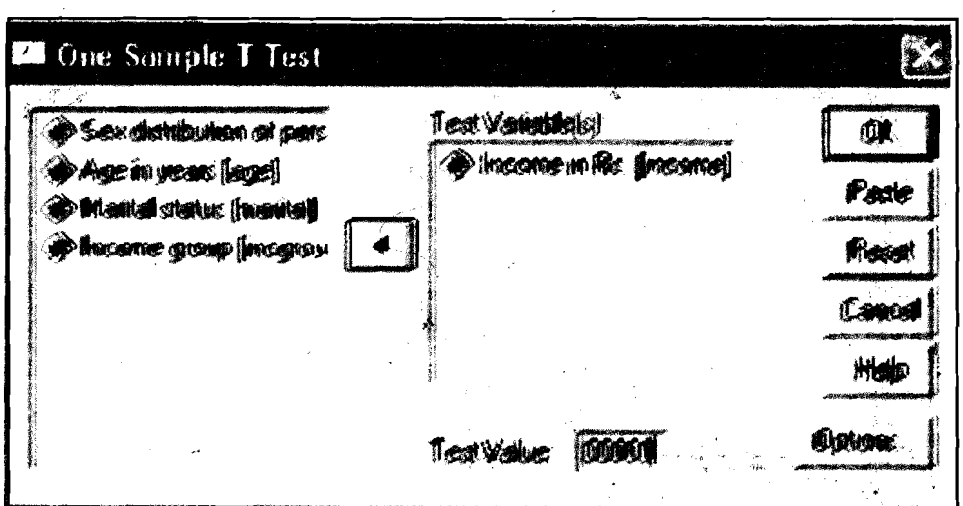


Figure 30.29 One Sample T Test

left side box. Shift this variable to the box under **Test Variable(s)** area.

3) Suppose you have selected Income Variable for the t-test and hypothesis takes the population mean income as Rs.200000. Type 200000 in the test box next to **Test Value**.

4) Click OK button. The following is the output you will see in Output Viewer Window.

T-Test

Table 30.22 One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Income in Rs.	30	234810.00	192833.232	35206.370

Table 30.23 One Sample Test with Test Value

	Test Value = 200000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Income in Rs.	.989	29	.331	34810.00	−37195.11	106815.11

Two sample T-Test

Sometimes, our data analysis focuses on two distinct groups within a single population or we may want to compare the two populations in terms of their respective means. The two-sample t-test for the equality of means will help you in this.

1) Select **Analyse→Compare Means→Independent-Samples T**

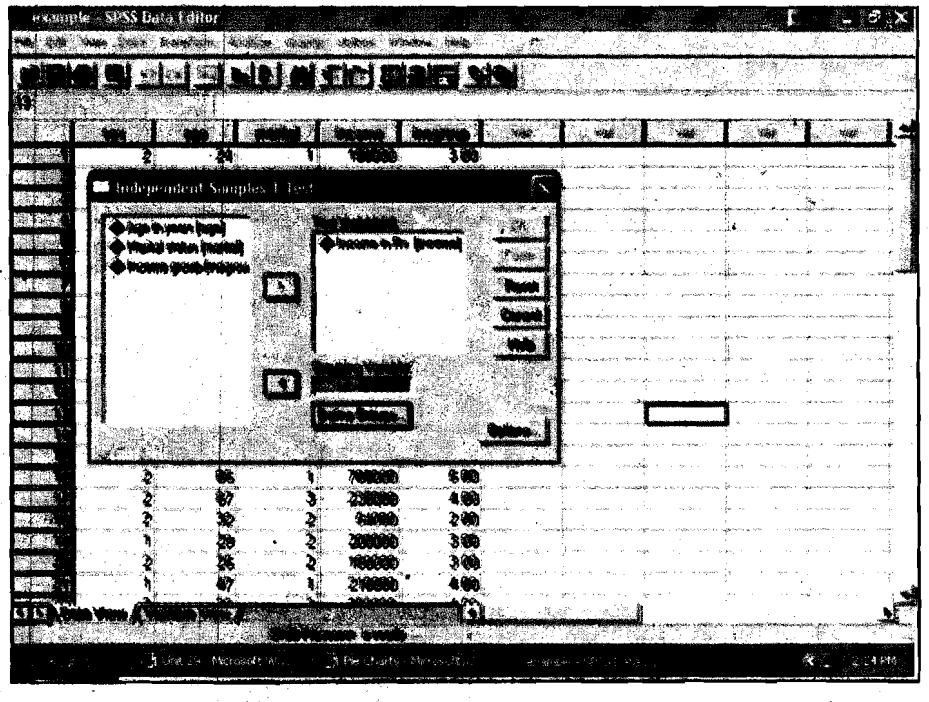


Figure 30.30 Example SPSS Data Editor: Income Variable

Test...command from the menu bar. The **Independent-Samples** T dialog box appears on the screen.

2) Click on the Variable name in the left side box that you want to include for the t-test. Suppose you want to include Income Variable, shift this variable to the right side box under **Test Variable(s)** using arrow key.

3) Click on the variable you want to group. Suppose you want to test the income mean difference between males and females, click on sex variable. Shift this variable to the box under **Grouping Variable** using arrow key.

4) Click on the **Define Groups**...button. The **Define Group** dialog box appears on the screen. Type '1' in the **Group 1** box. Type '2' in the **Group 2** box. Click **Continue** button to close the Define Groups dialog box.

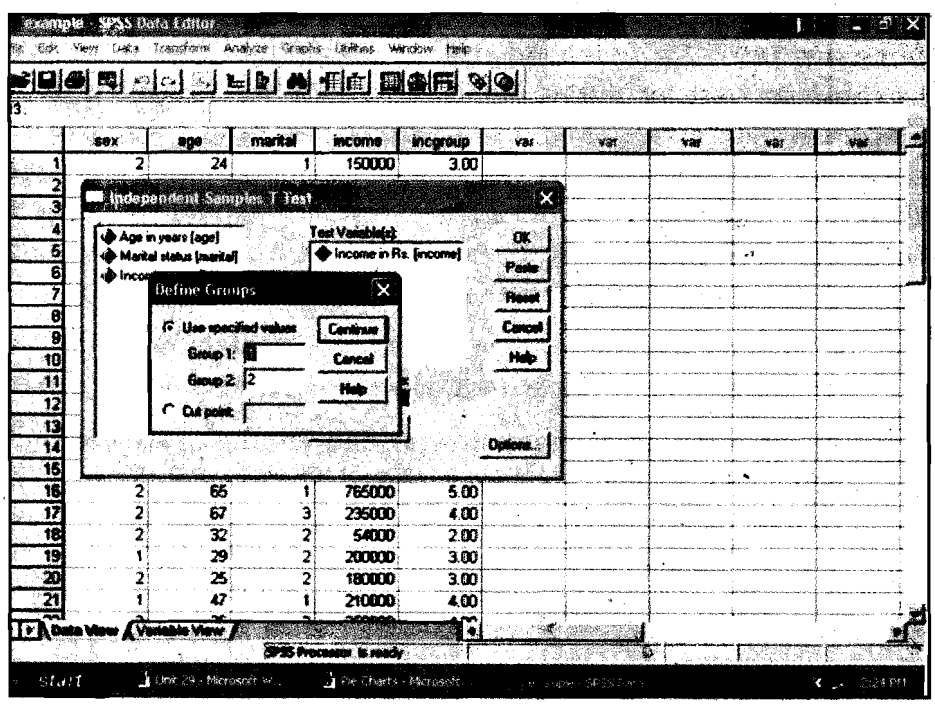


Figure 30.31 Example SPSS Data Editor: Grouping Variables

5) Click OK button to close the Independent-Samples T Test dialog box. The following table presents the output.

T-test

Table 30.24 Group Statistics

	Sex distribution of persons	N	Mean	Std. Deviation	Std. Error Mean
Income in Rs.	Male	14	262450.00	176610.187	47201.058
	Female	16	210625.00	208616.994	52154.248

Table 30.25 Independent Samples Test

			Income in Rs.	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		.118	
	Sig.		.734	
t-test for Equality of Means	t		.728	.737
	df		28	27.978
	Sig. (2-tailed)		.472	.467
	Mean Difference		51825.00	51825.00
	Std. Error Difference		71147.913	70342.061
	95% Confidence Interval of the Dfference	Lower	−93914.893	−92269.306
		Upper	197564.893	195919.306

Chi-square test© for independence

The chi-square test for testing the independence of attributes is for the categorical data arranged in a cross tabulation. The chi-square test appears as an option within the procedure for generating a cross-tabulation. The steps for chi-square test are repeat of steps for generating cross-tabulation with the addition of Expected counts under cells and chi-square under Statistics.

1) Select **Analyse→Descriptive Statistics→Crosstabs** command from the menu bar. The cross tabs dialog box appears on the screen.

2) Select a variable under **Row(s)** and another variable under **Column(s)**. Suppose you have selected Income level (a categorical variable with income levels low and high) variable under Row(s) and sex variable under Column(s).

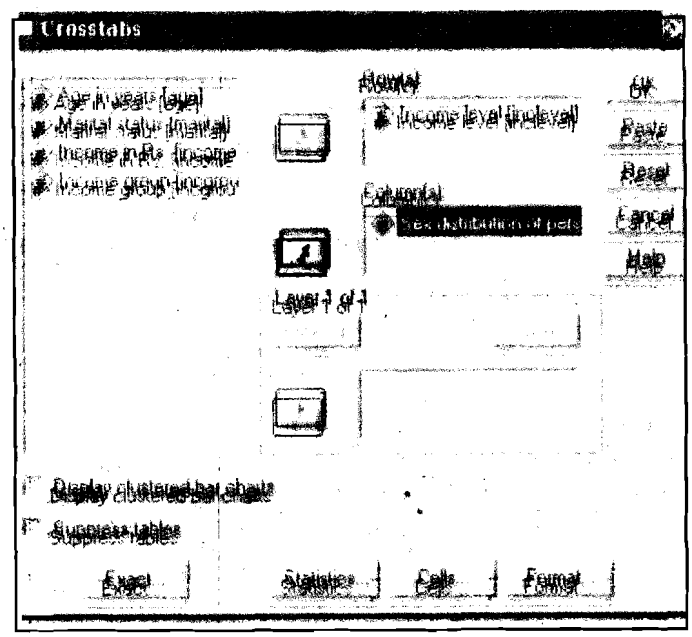


Figure 30.32 Crosstables: Income Levels

3) Click on the **Statistics**...button in the Cross tabs dialog box. The **Cross tabs:Statistics** dialog box appears on the screen. Check mark the Chi-square button. Click **Continue** button to close the Cross tabs: Statistics dialog box.

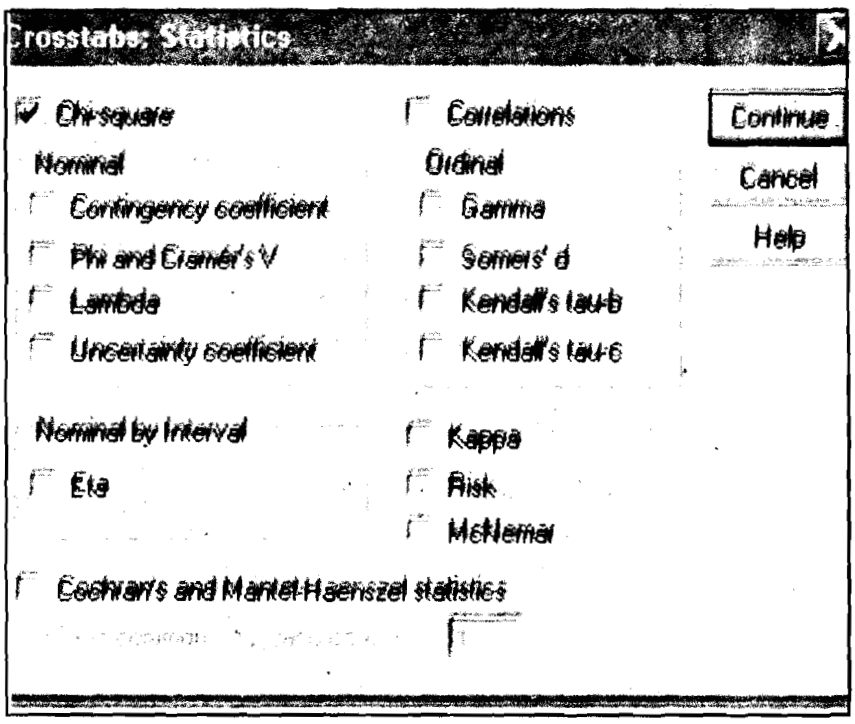


Figure 30.33 Crosstabs: Statistics

4) Click on the **Cells**...button in the Cross tabs dialog box. The **Crosstabs:Cell Display** dialog box appears on the screen. Check mark both Observed (if not already check marked) box and **Expected** box under Counts. Click **Continue** button to close **Cross tabs:Cell Display** dialog box.

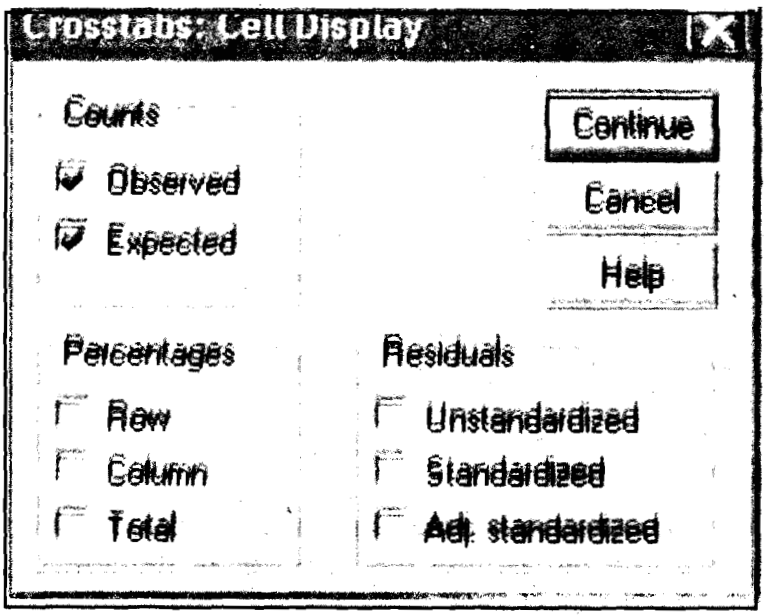


Figure 30.34 Crosstabs: Cell Display

5) Click OK button in the Cross tabs dialog box. The output generated is shown below.

Crosstabs

Table30.26 Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Income group * Sex distribution of persons	30	100.0%	0	.0%	30	100.0%

Table30.27
Income level * Sex distribution of persons Crosstabulation

			Sex distribution of persons		
			Male	Female	Total
Income level	Low	Count	4	8	12
		Expected Count	5.6	6.4	12.0
	High	Count	10	8	18
		Expected Count	8.4	9.6	18.0
Total		Count	14	16	30
		Expected Count	14.0	16.0	30.0

Table30.28 Chi-Square Tests

	Value	df	Asymp. Sig. (2-tailed)	Exact Sig. (2-tailed)	Exact Sig. (1-sided)
Pearson Chi-Square	1.429[b]	1	.232		
Continuity Correction[a]	.675	1	.411		
Likelihood Ratio	1.448	1	.229		
Fisher's Exact Test				.284	.206
Linear-by-Linear Association	1.381	1	.240		
N of Valid Cases	30				

a) Computed only for a 2X2 table
b) 0 cells (.0%) have expected countless than 5. The minimum expected count is 5.60.

Reflection and Action 30.3

Answer the following question on the basis of what you have just finished reading.

- What are the different views available in SPSS Data Editor?
- Explain when would you use each of these views?

30.8 Conclusion

In this unit you have learned the use of the SPSS Program to enter the data in a data file and use this data file for the analysis of data. You might have generated a data file using some other data base programs such as Excel. It is very easy to convert such data files

into a SPSS data file.

This unit provides an introduction to the SPSS. You can do a range of statistical analyses from simple cross tabulation to more complex statistical techniques, depending upon the individual researcher's requirement. However, we have tried to explain only simple commands and statistical tools, which are more popular in social research. We will leave it to the student to try and learn the full range of features in SPSS.

Further Reading

Nie, N. H., C.H. Hull, J. G. Jenkins, K. Steinbrenner and D. H. Bent 1979. *Statistical Package for the Social Sciences*. McGraw Hill: New York