UNIT 14 DATA MANAGEMENT, ANALYSIS AND PRESENTATION

Structure

14.1 Introduction

14.2 Objectives

14.3 Introduction to SPSS

14.4 Features of SPSS for Windows

14.5 Getting Started

14.6 Menu Commands and Sub-commands

14.7 Basic Steps in Data Analysis

14.8 Entering, Editing and Deleting Data

- 14.8.1 Entering Data
- 14.8.2 Inserting or Deleting Single Cases
- 14.8.3 Inserting or Deleting Single Variables

14.9 Importing Data into SPSS

14.10 Data File Management Functions

- 14.10.1 Merging Data Files
- 14.10.2 Aggregate Data
- 14.10.3 Split File
- 14.10.4 Select Cases

14.11 Running a Preliminary Analysis

- 14.11.1 Six Characteristics of a Dataset
- 14.11.2 Data Transformation
- 14.11.3 Graphical Presentation of Data

14.12 Understanding Relationship Between Variables: Data Analysis

- 14.12.1 Parametric Test
- 14.12.2 Non-Parametric Tests

14.13 SPSS Production Facility

14.14 JMP Statistical Analysis System (SAS)

14.15 NUDIST

14.16 Let Us Sum Up

14.1 INTRODUCTION

Statistical analysis and interpretation of data now forms an integral part of policymaking, strategic planning and related support services in the area of health and nutrition. Data analysis helps in the interpretation of data and takes a decision or answer the research question. Data analysis and interpretation have now taken center stage with the advent of the digital age. The purpose of analysing data is to obtain usable and useful information. Besides, a new discipline called 'Data Warehousing and Mining' has emerged in recent years, which focuses on the use and management of large and complex historical and contemporary database to extract information for policy makers and planners.

The methodological advances in quantitative analysis are also accompanied by a significant revolution in the computing power of the computer. The use of MS Excel software package that it is an electronic spreadsheet, about which you may recall reading in the Course 'Understanding Computer Application' (MFN-010) is one such example. You would certainly be familiar with this software and would know that it provides statistical, analytical and scientific functions. It has various features to offer namely fast calculations, what-if analysis, charts (graphs), automatic re-calculations and many more. Besides this package, the most popular software package of data analysis used today is the statistical package for the social sciences (SPSS).

SPSS is a comprehensive and flexible statistical analysis and data management system. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and conduct complex statistical analyses.

SPSS offers a wide range of methods, graphs and charts. In SPSS, the output is kept separate from the data itself. It stores the data in a separate file.

In this unit we will focus on SPSS fundamentals and the use of its statistical components. We shall also look at the several statistical techniques for qualitative and quantitative data analysis and discuss situations in which you would use each of these techniques, the assumptions made by each method, how to set-up analysis using SPSS,as well as, how to interpret the results.

We assume that you have already acquired basic knowledge of statistical methods (through this course) and the use of computers (through the MFN-010 Course). Given these skills, this unit will help you get acquainted with the statistical capabilities of SPSS for Windows. We will also introduce in this unit, JMP StatisticalAnalysis System (SAS) software for quantitative data analysis. For qualitative dataanalysis, we will introduce software called NUDIST.

14.2 OBJECTIVES

After studying this unit, you will be able to:

- explain/describe the main features of the SPSS;
- write about, as well as, use the data management operations and techniques of analysis using SPSS;
- acquire skill in the use of SPSS for basic statistical analysis with a special focus on the measures of central tendency, dispersion, correlation and regression etc.;and
- Present the data and the SPSS results graphically.

14.3 INTRODUCTION TO SPSS

SPSS means "**S**tatistical **P**ackage for the **S**ocial **S**ciences" and was first launched in 1968. Since SPSS was acquired by IBM in 2009, it's officially known as IBM SPSS Statistics but most users still just refer to it as "SPSS". SPSS is a user-friendly program that facilitates data management and statistical analyses. It is a statistical tool used by researchers on a large scale to perform quantitative

analysis of the data. SPSS has a very easy to interact with interface which is in the form of a spreadsheet. SPSS can read files in either of the formats such as MS-Excel, Plain text files (.txt or .csv) or Relational databases(SQL).

Once the data has been collected, the first step is to look at it in a variety of ways. While there are many specialized software application packages for different types of data analysis (relating to scientific, commercial and financial problems), a researcher is often faced with a situation where the general treatment and standard statistical analysis of the quantitative data is required. SPSS is one-such package that is often used by researchers and analysts for data management and exploring it before attempting a detailed statistical analysis. It is a preferred choice for research analysis due to its easy-to-use interface and comprehensive range of data manipulation and analytical tools.

Suppose, you are interested in knowing the attitude of students towards distance education and for that you have administered a data collection instrument (commonly known as questionnaire) to some students. Now you want to process and analyze the data. Till recently, data were processed manually and it was indeed a cumbersome process. Fortunately, now we live in an age when high-speed computers can do the job of processing and analysis of data in a very short period of time and of course without errors. What you have to do is you have to learn some fundamental concepts used in this programme. Now you can sit at the computer and process and analyse the data that you have collected by administering a questionnaire. In fact, you will find it helpful and interesting to keep the SPSS Application guide nearby you while you process and analyze your data.

Here to help you work with the SPSS, some general features are highlighted next.

14.4 FEATURES OF SPSS FOR WINDOWS

SPSS is one of the leading desktop statistical packages. It is an ideal companion to the database and spreadsheet, combining many of their features, as well as, adding its own specialized functions. SPSS for Windows is available, as a base module and a number of optional add-on enhancements are also available. Some versions present SPSS as an integrated package including the base and some important add-on modules.

It's available in four license options: Base Edition, Standard Edition, Professional Edition and Premium Edition.

SPSS Base Edition provides core statistical and graphics capabilities, advanced statistical procedures, Nonlinear regression and Customized tables to analyze and report on numerical and categorical data (not available in Statistics Standard Edition for Academia).

SPSS Statistics Standard edition offers the core statistical procedures for business managers and analysts need to address fundamental business and research questions. This software provides tools that allow users to quickly view data, formulate hypotheses for additional testing, and carry out procedures to clarify relationships between variables, create clusters, identify trends and make predictions.

SPSS Professional Statistics edition goes beyond the core statistical capabilities offered in the Standard Edition to address issues of data quality, data complexity, and automation and forecasting. It is designed for users who perform many types of in-depth and non-standard analyses and who need to save time by automating data preparation tasks.

SPSS Statistics Premium Edition helps data analysts, planners, forecasters, survey researchers, program evaluators and database marketers – among others – to easily accomplish tasks at every phase of the analytical process. It includes a broad array of fully integrated Statistics capabilities and related products for specialized analytical tasks across the enterprise. The software will improve productivity significantly and help achieve superior results for specific projects and business goals.

SPSS creates a high quality presentation – quality tabular reports includings tub and banner tables and display of multiple response data sets. The new features include pivot tables, a valuable tool for presentation of selected analytical outputtables.

As of 2019, SPSS Statistics Version 26 is the most recent version of the SPSS Statistics traditional software license. You can consult the SPSS World Wide Web site for the latest developments and additions to the computing power of SPSS. Technical support is also available to the registered users at the SPSS site. The SPSS Web site is http://www.spss.com. Select white papers on SPSS applications in major disciplines are also available on this site.

This unit discusses some of the commonly used data, management techniques and statistical procedures using SPSS 26 version since new features are added almost daily, you are advised to check for these details on the currently installedversion of SPSS on your computer and also consult the user manuals before undertaking complex type of data analysis. The on-line help is also available. There may be some procedures and syntax-related changes from one version to another.

With this basic knowledge let us get acquainted with the SPSS.

14.5 GETTING STARTED

To start SPSS, click on the Start button and place the pointer on Programs, a sub-menu of Programs is displayed. In this menu, click on the IBM SPSS Statistics and then IBM SPSS Statistics to open the SPSS program.

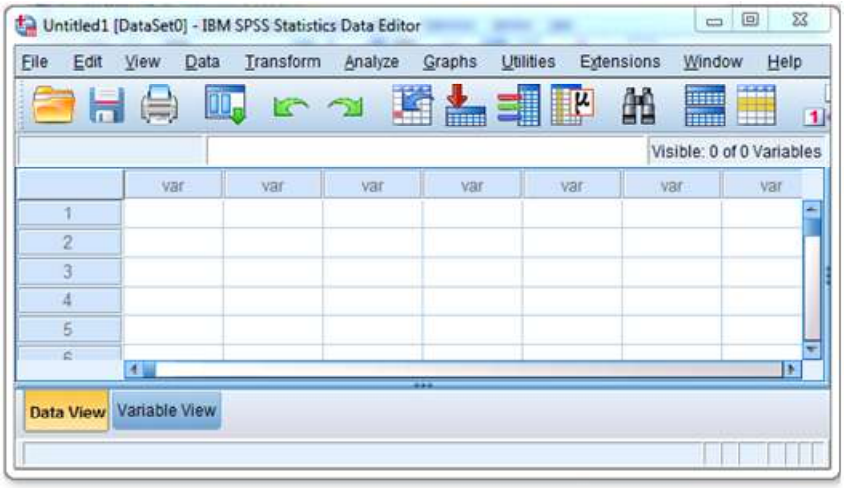

Fig.14.1: Data Editor Window (Data View)

A new blank Data Editor Window is displayed each time you start SPSS as shown in Figure 14.1.

In SPSS, three windows are commonly used i.e. Data Editor Window, Syntax Editor Window and Output Viewer Window.

Data Editor

The Data Editor window has two views (i.e. Data View and Variable View) that can be selected from the lower left hand side of the screen. Data View is where you see the data you are using. VariableView is where you can see additional information about your data.The data in the Data Editor is saved in a file with the extension .sav.Data Editor Window is the only window that's always open when you run SPSS.

SPSS Data View

SPSS Data View, you can see your actual data(see Figure 1). The Data View is a Spreadsheet which contains rows and columns, The Data can be entered in the Data View Sheet either manually or the data can be imported from the data file. SPSS can read the data file in any of the one format, which can be Excel, Plain text files or relational (SQL) Databases.

SPSS Variable View

In variable view, you can define the Variable of the Data that you have. The Variable View Consists of the Column heading such as Name, Type, Width, Decimals, Label, Value, Missing, Align and Measure as shown in Figure 14.2.

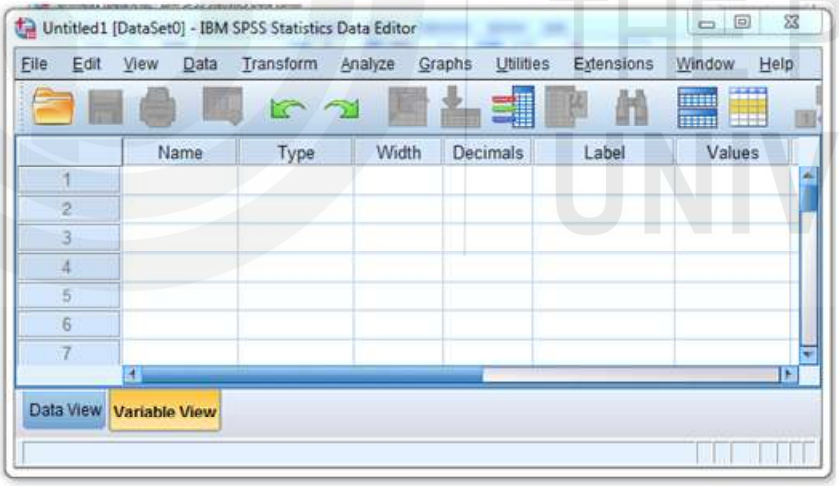

Fig.14.2: Data Editor Window (VariableView)

Output Viewer

When you perform any command in SPSS, the Output Viewer window opens automatically and displays a log of the actions taken and the associated output. Primarily, the Output Viewer is where the results of statistical analysis are shown, but any command invoked through the drop-down menus or syntax will be printed to the Output Viewer. This includes opening, closing, or saving a data file. If an Output Viewer window is not open when a command is run, a new Output Viewer window will automatically be created. The Output Viewer window is divided into two sections, or frames. The left frame contains an outline of the content in

the Output Viewer. The right frame contains the actual output. An Output Viewer window can be saved as a viewer file (*.spv).

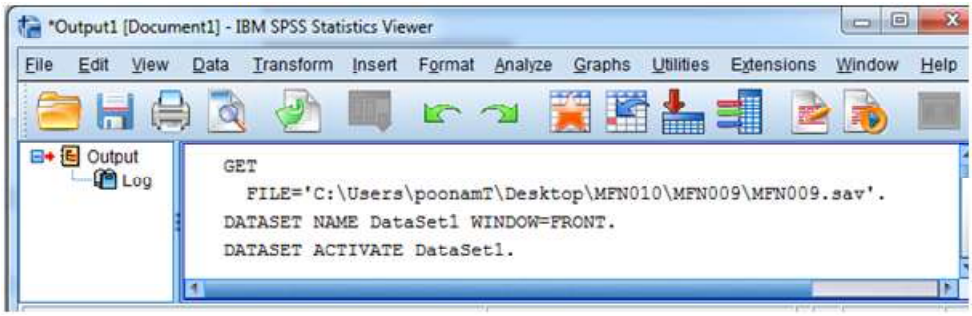

Fig.14.3: Output Viewer Window

Syntax Editor

SPSS syntax editor is a programming language unique to SPSS that can be used as an alternative to the drop-down menus for data manipulation and statistical analyses. The Syntax Editor window is where you can write, debug and execute SPSS syntax. You can open a new Syntax Editor window by using File > New > Syntax option.The right panel of the Syntax Editor window is where your syntax is entered. The left panel of the Syntax Editor window shows an outline of the commands in your syntax, and can be used to navigate within your code

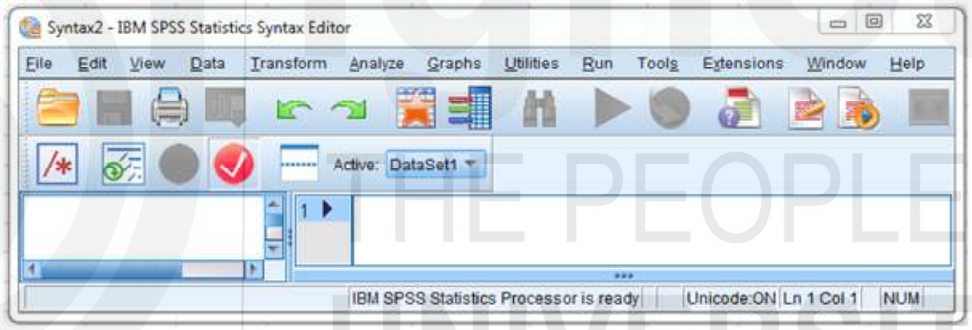

Fig.14.4: Syntax Editor Window

Script Editor

This facility is normally used by the advanced users. It offers fully featured programming environment that uses the Sax BASIC language and includes a Script Editor, Object Browser, Debugging features and context sensitive help. Scripting allows you to automate tasks in SPSS including:

- Automatically customizing output
- Open and save data files
- Display and manipulate SPSS dialog boxes
- Run data transformation and statistical procedures using SPSS command Syntax
- Export charts as graphic files in a variety of formats.

The present module will not go into the details of the advanced features of SPSS including scripting.

Exiting SPSS

Make sure that all SPSS and other files are saved before quitting the program. Youshould exit the software by shutting off the program by selecting Exit SPSS command from the file menu of the SPSS Data Editor window. In case of unsaved files, the SPSS will prompt you to save or discard the changes in the file.

Saving data and other files

Many types of file can be saved using 'save' or 'save as' command. Various types of file used in SPSS are: Data, -Syntax, Chart or Output. Files from spreadsheet or other databases can also be imported by following the appropriate procedure. Similarly, an SPSS file can be saved as a spreadsheet or in Dbase format. Select the appropriate save type command and save the file. The SPSS data files are saved with. sav as the secondary name. Though SPSS files could be given any name, the use of reserved words and symbols is to be avoided in all types of file names.

Printing of data and output files

The contents of SPSS data files, Output Navigator files and Syntax Files can be printed-using the standard 'Print' Command. The SPSS uses the default printer for printing. In the case of network printers, an appropriate printer should be selected for printing the output.

14.6 MENU COMMANDS AND SUB-COMMANDS

Most of the commands can be executed by making appropriate selections from themenu bar. Some additional commands and procedures are available only through the Syntax Window. Since SPSS is menu driven, each Window has its own menu bar. While some of the menu bars are common, the others are specific to a particular, type of Window. The menu and sub-menus of the Data Editor window are given below:

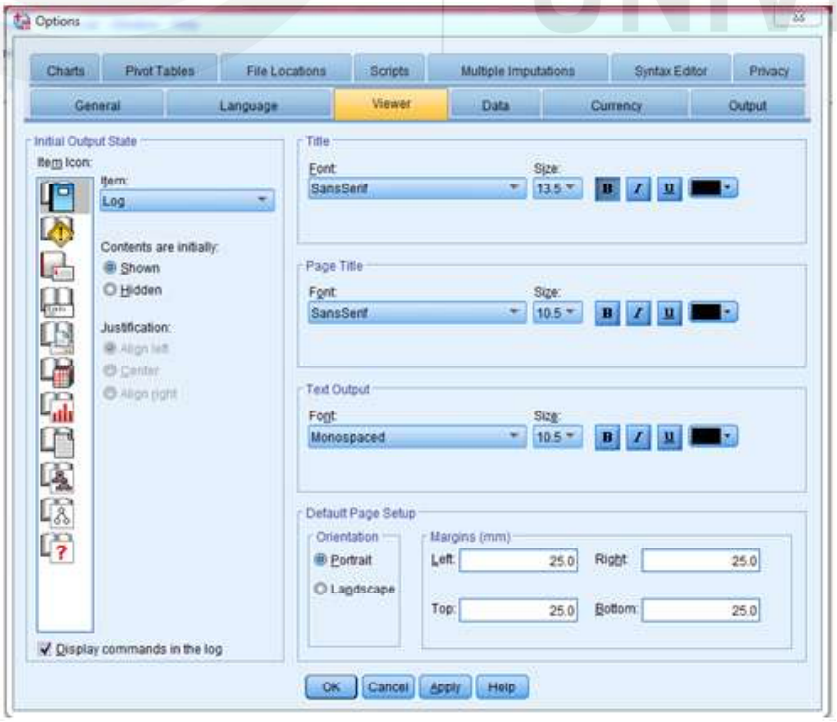

Fig.14.5: SPSS Options

File

File includes all of the options you typically use in other programs, such as open, save, exit. This menu is also used to import data created in other formats like database, excel, text files etc. Print control functions like page setup, printer setup and associated functions.

Edit

These functions are similar to those available in general packages.These include undo, redo, cut, copy, paste, copy variable, find, findand replace, insert variable and insert cases. Option setting for the SPSS is controlled through Edit menu.

The SPSS provides a facility for setting up of the user defined options. Use the Edit menu and then select Options. The following types of optional setting are allowed in SPSS as illustrated in Figure 14.5. Make the appropriate changes to set the options according to your choice.

View

View allows you to select which toolbars you want to show, select font size, add or remove the gridlines that separate each piece of data, and to select whether or not to display your raw data or the data labels.

Data

This is a very important menu as far as management of the data is concerned. Variable definition, identify duplicate cases, transposing templates, aggregating and merging of data files, splitting data filesfor specific analysis are some important commands in Data Menu.

Transform

Compute new variables, recede, random number generation, ranking, time series data transformation, count and missing value analysis are undertaken using Transform Command.

Analyze

As the name implies, analyze Menu incorporates statistical procedures,frequency distribution, cross-tabulations, comparison of means,correlation, simple and multiple regression, Log linear regression, discriminate analysis, factor analysis, non-parametric test and time series analysis are undertaken using analyze menu.

Graphs

Graphs include the commands to create various types of graphs including box plots, histograms, line graphs, and bar charts.

Utilities

It is useful for creating customised and automated analyses.

Window

Window can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax).

Help has many useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide. Using topics, you can use the index option to type in any key word and get a list of options, or you can view the categories and subcategories available under contents. This is an excellent tool and can be used to troubleshoot most problems.

The **Icons** directly under the Menu bar provide shortcuts to many common commands that are available in specific menus. Take a moment to review these as well.

With this basic knowledge about commands and sub-command, now let us learn about the basic steps in data analysis.

14.7 BASIC STEPS IN DATA ANALYSIS

There are some basic steps involved in data analysis using SPSS. These are given below:

Get your data into SPSS

You can bring your data into SPSS in thefollowing ways:

- Enter data directly in the Data Editor
- Open a previously saved SPSS Statistics data file
- Read a spreadsheet, database or text data file into SPSS data editor

Select a procedure

Select a procedure from the menus to calculate statistics or to create a chart.

Select the variables for the analysis

The variables in the data file are displayed in a dialog box for the procedure.

Run the procedure

Run the procedure and look at the results. Results are displayed in the Output Viewer.

So then the basic steps involved in data analysis are clear. Before analysis we needto define, edit and enter data. Let us get to know about this next.

14.8 ENTERING, EDITING AND DELETING DATA

There are many options for creating SPSS data files. The data can either be directly entered through Data Editor or imported from spreadsheet, ASCII file and other RDBMS packages like oracle and Access. Let us understand how to start, define, edit and enter data in the SPSS.

When you start SPSSsoftware, the Data Editor opensautomatically. The Data Editor provides a spreadsheet for entering and editing dataand creating data files. (See the Figure 14.6).

In Data Editor Window, you will see that each of the columns is labeled as “var.” The column names will represent the variables that you enter in your dataset. You

will also see that each row is labeled with a number (i.e. "1", "2" and so on). The rows will represent cases that will be a part of your dataset. When you enter values for your data in the spreadsheet cells, each value will correspond to a specific variable (column) and a specific case (row).

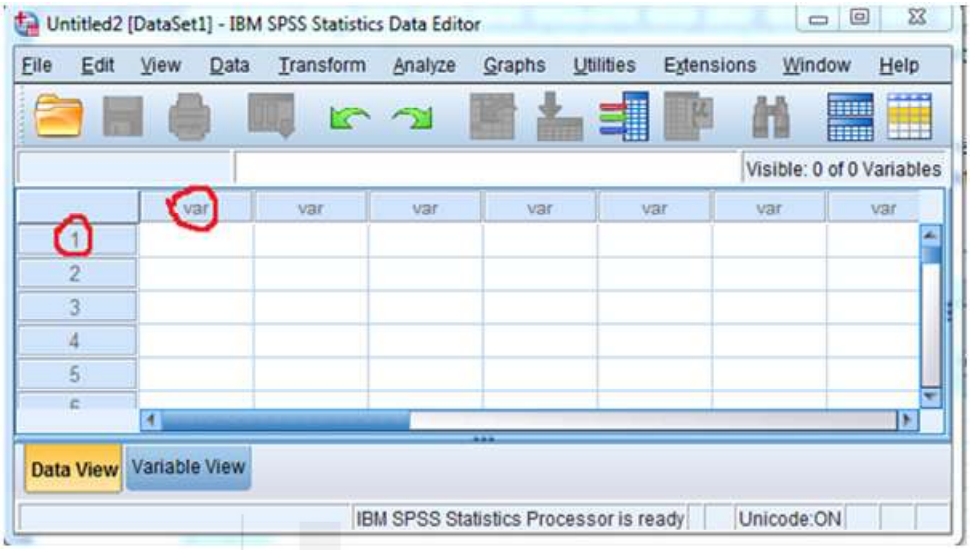

Fig.14.6: Data Editor

14.8.1 Entering Data

You can follow these steps to enter data:

· Click the **Variable View** tab. Type the name for your first variable under the Name column. You can also enter other information about the variable, such as the type (the default is "numeric"), width, decimals, label, etc. Type the name for each variable that you want to include in your dataset. For example, we will write' Roll_No' under the Name column and select Numeric data type for Type column as shown in following Figure 14.7.

Fig.14.7: Entering Data (1) in Variable view

- Click the **Data View** tab. Any variable names that you entered in Variable View will now be included in the columns (one variable name per column). You can see that Roll_No appears in the first column in this example. (see the Figure 14.8).

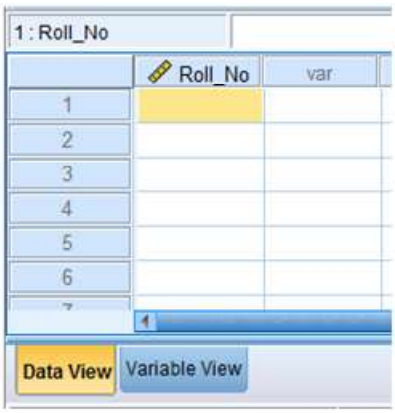

Fig.14.8: Entering Data (2) in Data view

- Now you can enter values for each case. In this example, cases represent students. For each student, enter a value for their name in the cell that corresponds to the appropriate row and column. For example, the first student information should appear in the first row, under the variable column 'Roll_No". In this example, you have entered 5 cases in your dataset as per following Figure 14.9.

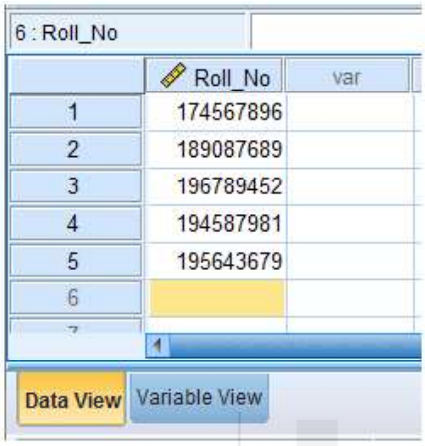

Fig.14.9: Enteing Data (3) in Data view

- Repeat these steps for each variable that you will include in your dataset like the following Figure 14.10.

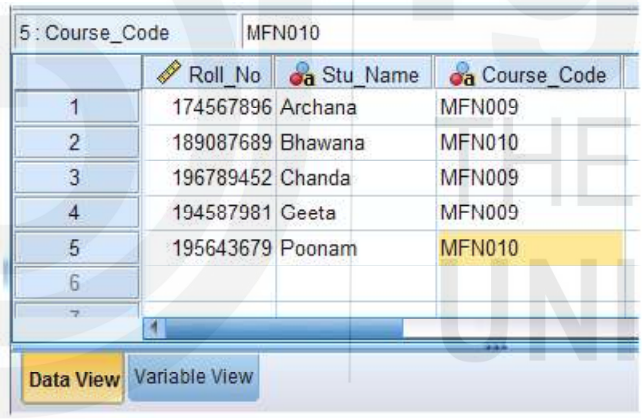

Fig.14.10: Entering Data (4) in Data view

- Now save the data.

To **save data,** select save optionfrom the File menu and click on save.

Because these data have not been saved previously you will see a dialogue box prompting you to enter a file name. Type in the name "attitude" and click OKbutton. SPSS will then save the data to this file. (SPSS will automatically attach the.sav extension if you do not type it in.)

14.8.2 Inserting or Deleting Single Cases

You can also add new cases or delete the existing cases from your dataset. For this, you can do the following:

To insert a new case into a dataset:

- In Data View, Click on **Edit** menu and select **Insert Cases OR**
- Right-click on a row and select Insert Cases from the menu

- A new blank row will appear above the row you selected.
- Type in the values for each variable in the new case.

To delete an existing case from a dataset:

- In the Data View tab, click the case number (row) that you want to delete. This will highlight the row for the case you selected.
- Press Delete on your keyboard or right-click on the case number and select "Clear". This will remove the entire row from the dataset.

14.8.3 Inserting or Deleting Single Variables

Sometimes you may need to add new variables or delete existing variables from your dataset. You can do this using following steps:

To insert a new variable into a dataset:

- In the Data View, click the name of the column to the right of where you want your new variable to be inserted
- click on **Edit** menu and select **Insert Variable** OR
- Right-click an existing variable name and click Insert Variable
- A new blank column will appear to the left of the column or cell you selected. (see Figure 14.11).

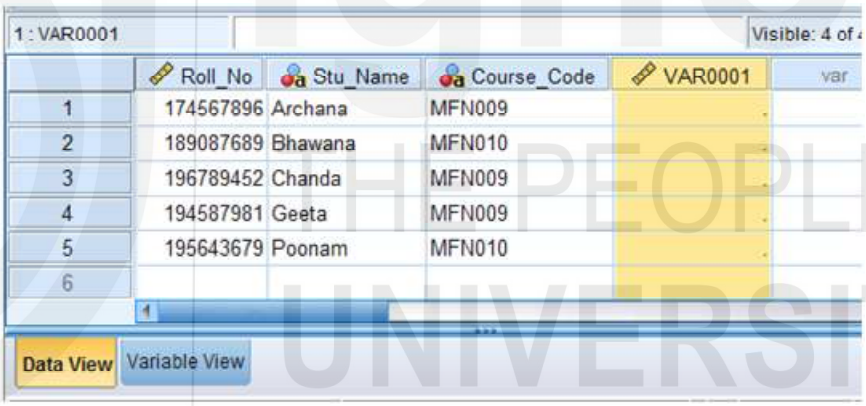

Fig.14.11: Inserting a new variable

New variables will be given a generic name (e.g. VAR0001). You can enter a new name for the variable on the Variable View tab.

To delete an existing variable from a dataset:

- In the Data View tab, click the column name (variable) that you wish to delete. This will highlight the variable column.
- Press Delete on your keyboard, or right-click on the selected variable and click "Clear." The variable and associated values will be removed. OR
- You can also delete a variable through the Variable View window.

14.9 IMPORTING DATA INTO SPSS

If you are already created data in ".sav" file format,you can simply open that file to begin working with your data in SPSS. However, if you have data stored in

other types of filessuch as Excel files and text files then you will need to import data from other source and then save it in the SPSS file format. This section will cover how to import data from Excel files.

To import data from an Excel spreadsheet into SPSS, first make sure your Excel spreadsheet is formatted according to SPSS criteria:

- The spreadsheet should have a single row of variable names across the top of the spreadsheet in the first row.
- The data should begin in the first column, second row (beneath the variable names row) of the spreadsheet.
- Variable names should include ordinary letters, numbers, and underscores (e.g., Name, St_Name, Date_1) and not include special characters (e.g., "Student Name" would not be a valid variable name because it contains a space).
- Anything like graphs, Pivot Tables, labels which are not part of data should be removed.

Here is an example of properly formatted data looks like in Excel 2010 (see Figure 14.12):

	A	B	C	D	E	F	G
1	Roll_No	Name	Course	MARKS	ASGN1	LAB1	LAB2
2	168967780	Archana	MFN009	45	67	56	56
3	195687896	Bhawna	MFN009	49	56	40	45
4	189878675	Chanda	MFN009	56	78	55	55
5	184069454	Dolly	MFN009	47	55	61	61
6	178260233	Geeta	MFN010	65	45	45	46
7	172451012	Poonam	MFN011	66	58	46	60
8	166641791	Sunita	MFN012	55	60	50	50

Fig.14.12: Sample formatted data in Excel S/W

Now, you can import data from an Excel File using the following steps:

- Click on File menu (see Figure 14.13).
- Select Import Data and click on Excel.

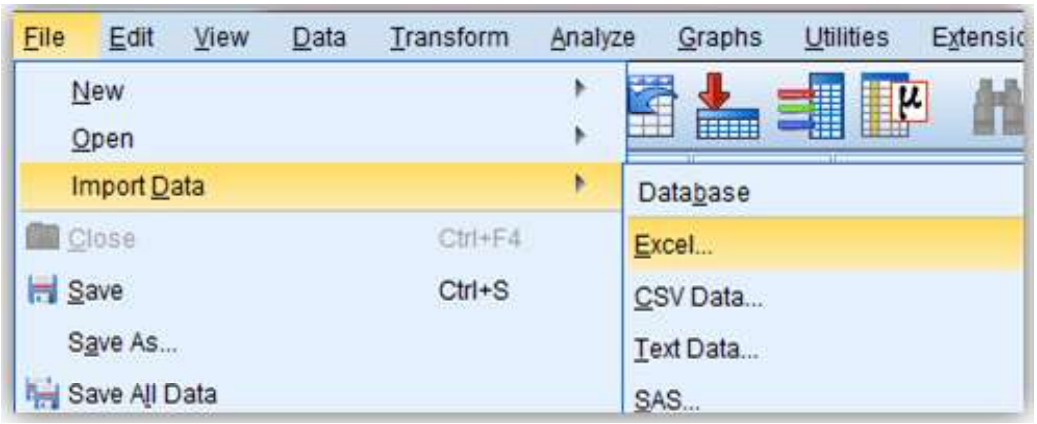

Fig.14.13: File menu

· The Open Data window will appear (see Figure 14.14).

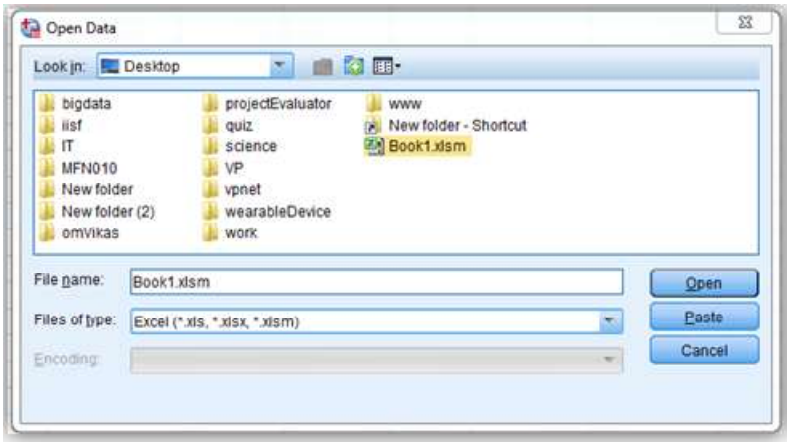

Fig.14.14: Open Data Window

- Select your excel file and press open button.
- Now, Read Excel File window will appear (see Figure 14.15)

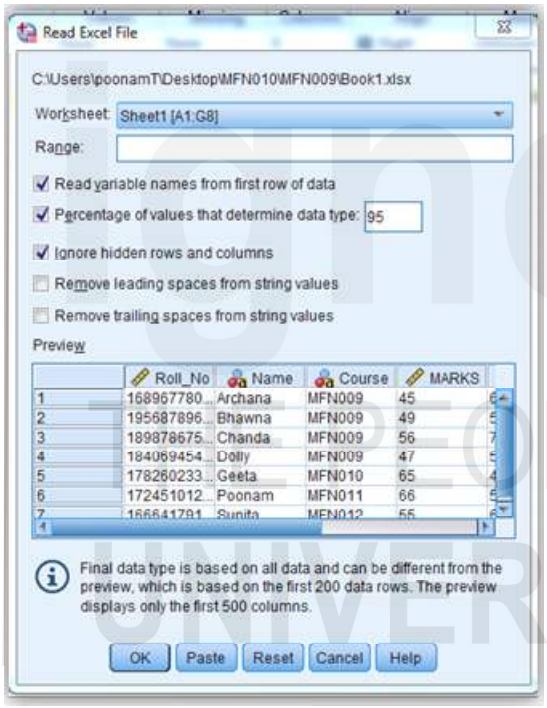

Fig.14.15: Read Excel File window

- Click OK.

Now the data will appear in SPSS. Here is an example of how sample data appear in SPSS once the data have been imported (see Figure 14.16):

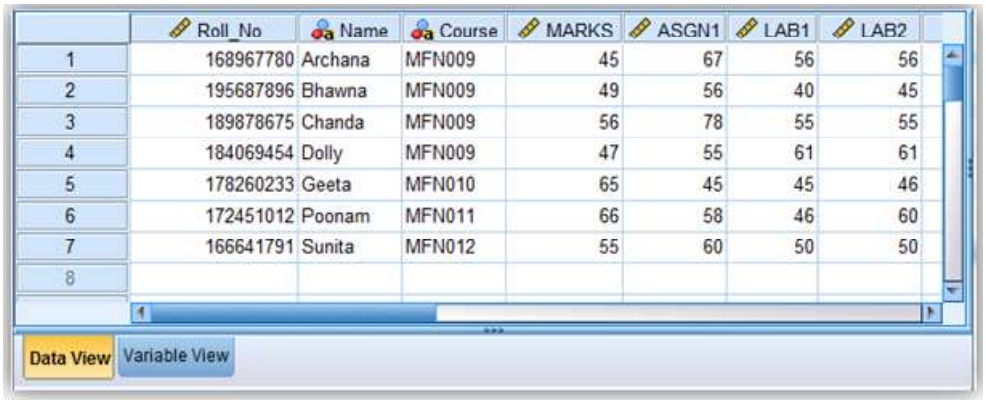

Fig.14.16: sample data appear in SPSS

14.10 DATA FILE MANAGEMENT FUNCTIONS

SPSS is very flexible as far as management of data files is concerned. While only onefile can be opened for analysis at a time, the SPSS provides flexibility in merging multiple data files with the same structure into one single data file, merging files to add new variables, partially select the cases for analysis, and make groups of data based on certain characteristics and use different weights for different variables. Some of these functions are discussed below. Groups of data can also be defined to facilitate the analysis of the most commonly referred variables (see utilities and data commands).

14.10.1 Merging Data Files

Researchers are often faced with a situation where data from different files are to be merged or a limited number of variables from large and complex data files are required. The following types of facility are available for merging files using SPSS.

Adding cases

This option is used when the data from two fileshaving the same variables are to be combined. For example, you may record the same information for students in different study centres in India and abroad. The datacan be merged to create a centralized database by using Add cases command.

Adding variables

Adding variables is useful when two data files contain the information about same case but on different variables. For example, the teacher's database may contain two files, one having the educational qualifications and the other having the names of the courses taught. Both the files could becombined to analyze the variables available in them. The data on a key and uniquevariable from both the files can be combined easily. The key variables must have thesame name in both the data files. Both the data files should be sorted on the common key variable.

Consider the following example for merging files:

For merging files, create 2 data files taking sample dataset which are given below:

- Take Sample dataset 1 (see figure 14.17)

1 : Roll_No 174567896 Visible:

	Roll_No	Course	MARKS	ASGN1	LAB1	LAB2
1	174567896	MFN009	45	67	56	56
2	189087689	MFN010	49	56	40	45
3	194587981	MFN009	47	55	61	61
4	195643679	MFN010	65	45	45	46
5	196789452	MFN009	56	78	55	55
6						
7						

Data View Variable View

Fig.14.17: Sample dataset 1

Take Sample dataset 2 (see figure 14.18)

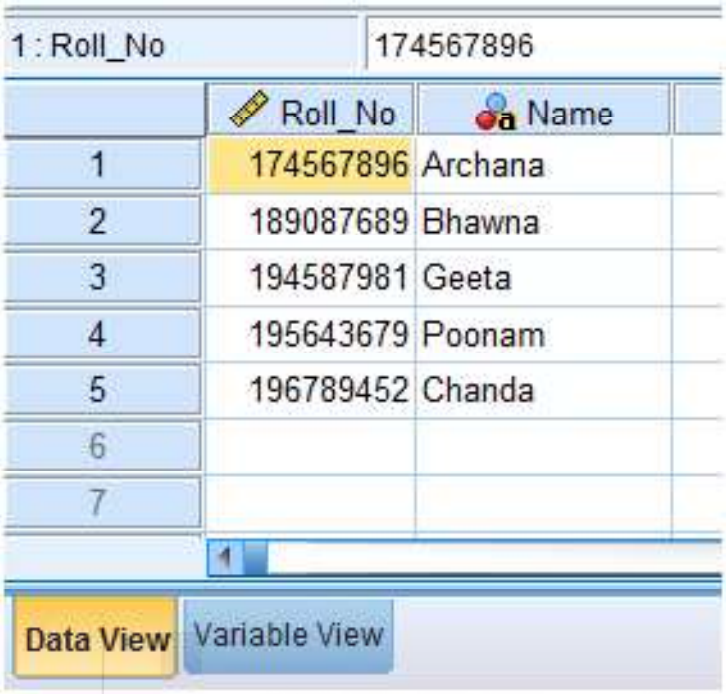

Fig.14.18: Sample dataset 2

- Open at least one of the data files that you want to merge. If you have multiple datasets open, make one of the datasets that you want to merge the active dataset (for e.g. Dataset 1)
- Select **Data**>**Merge Files**From the menu and then select **Add Variables**
- Select the dataset or external data file to merge with the active dataset. (see Figure 14.19).

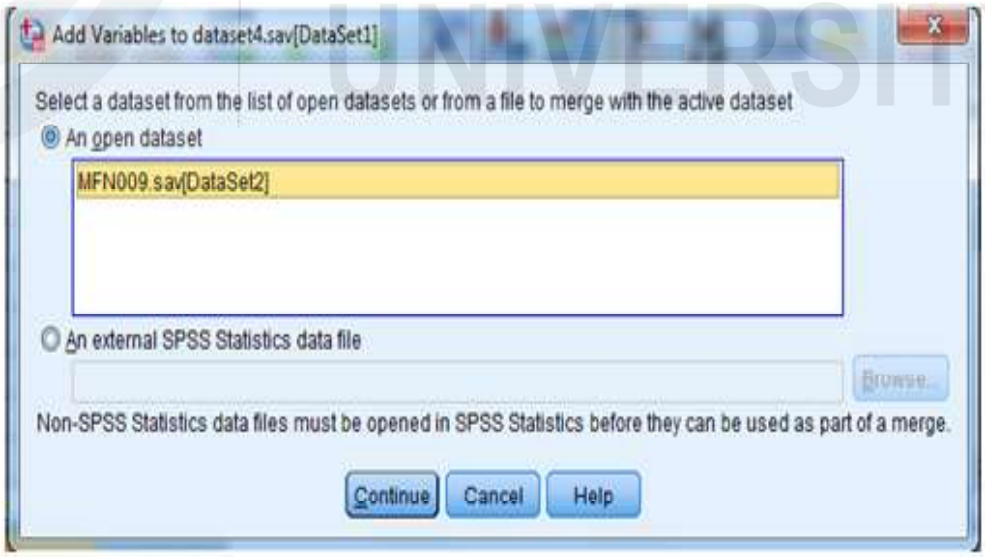

Fig.14.19: Merge File (1): Add data file

- Click on Continue button
- Now open a new window for Add Variable from other Dataset (See Figure 14.20)

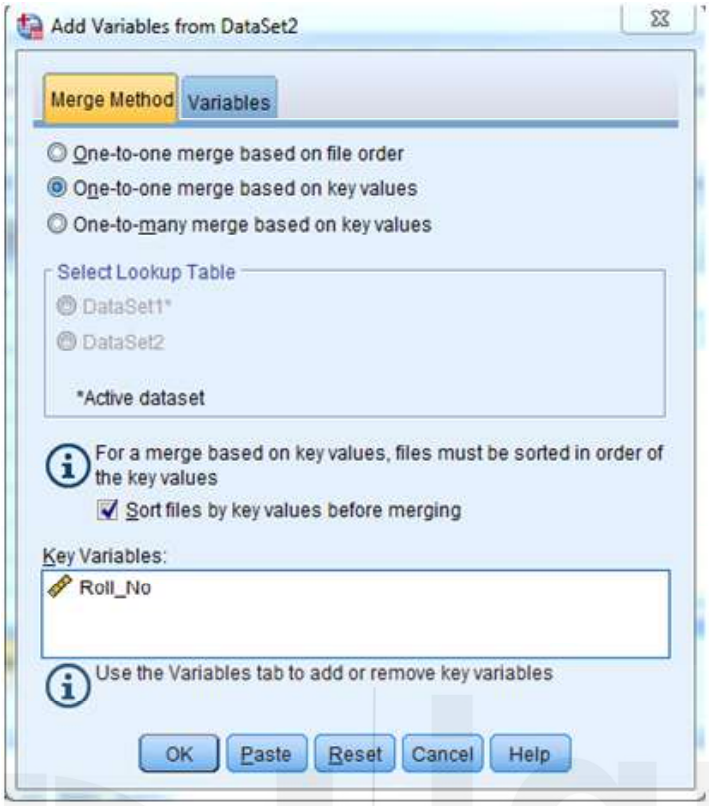

Fig.14.20: Merge File (2): Add Variable from other Dataset

- On Merge Method tab, select Match cases on key variables in sorted files. (In this example, select one-to-one merge based on key values)
- The key variables must exist in both the active dataset and the other dataset.

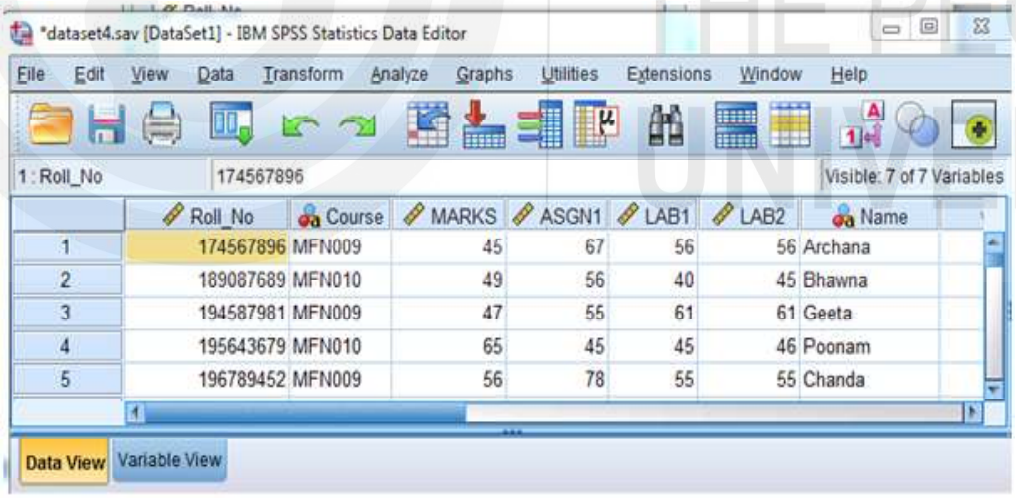

Fig.14.21: Merging files: New dataset

Click OK button.

- After merging files, the new dataset is displayed as shown in following figure 14.21.

14.10.2 Aggregate Data

Aggregate Data command combines groups of cases in the active dataset into a single summary case and creates a new aggregated data file. Cases are aggregated,

based on the value of *zero* or more break (grouping) variables. If no break variables are specified, then the entire dataset is a single break group. The new (aggregated) file contains one record for each group. The aggregate file could be saved with a specific name to be provided by you the user. Otherwise, the default name is aggregate.sav. For example, the data on learners, achievement could be aggregated by state andregion.

A number of aggregate functions are available in the SPSS. These include sum, mean, number of cases, maximum value, minimum value, standard deviation, first andthe last value. Other Summary functions include percentage and fractions below and above a particular cut-off user-defined value.

Break Variables

Cases are grouped together based on the values of the break variables. Each unique combination of break variable values defines a group. When creating a new, aggregated data file, all break variables are saved in the new file with their existing names and dictionary information. The break variable, if specified, can be either numeric or string.

Aggregated Variables

Source variables are used with aggregate functions to create new aggregate variables. The aggregate variable name is followed by an optional variable label, the name of the aggregate function, and the source variable name in parentheses. (See the Figure 14.22)

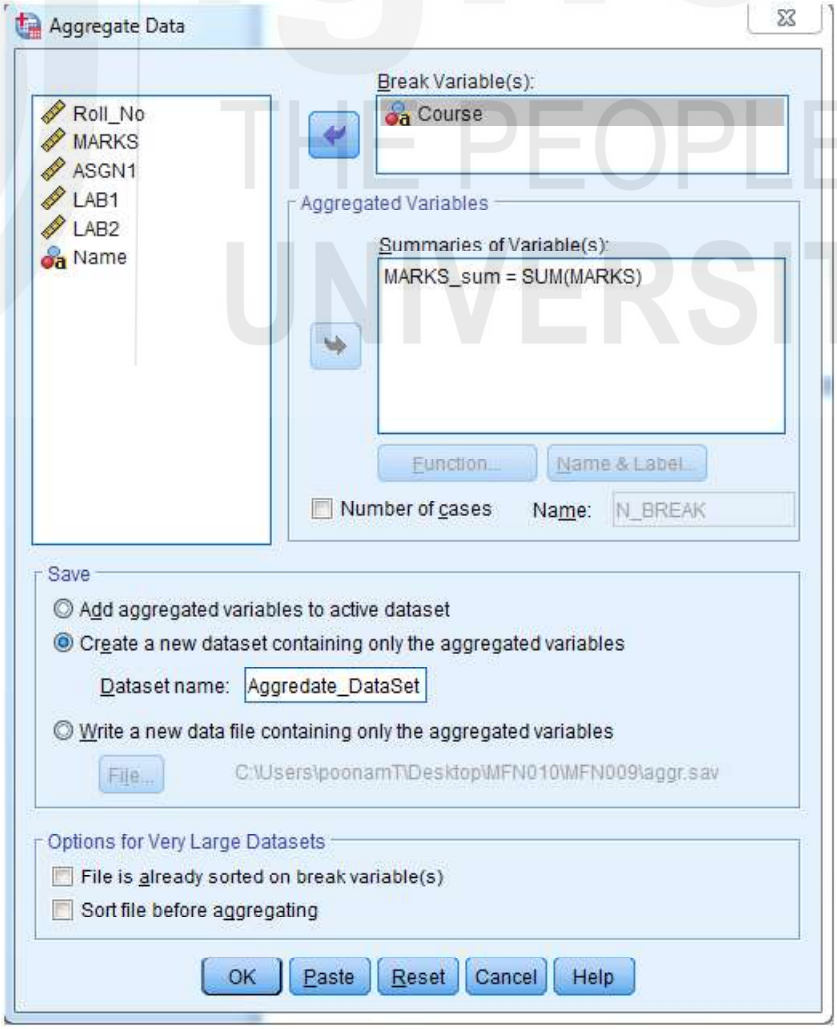

Fig.14.22: Aggregate Data

Aggregate functions include:

- Summary functions for numeric variables, including mean, median, standard deviation, and sum
- Number of cases, including unweighted, weighted, nonmissing, and missing
- Percentage, fraction or count of values above or below a specified value
- Percentage, fraction, or count of values inside or outside of a specified range

Consider the following example for Aggregate Data, you can follow these steps:

- Create a data file like the dataset defined in Figure 14.17.
- Select Data from menu bar and then select Aggregate command
- A dialog box will appear in which you can define break and aggregate variables on the basis of aggregation will be done.

You can see data variable in the left side of the dialog box. You can define Course as break variable and Marks as aggregate variable.

- Click Function button.
- Now, Aggregate Function dialog box will be open. (see Figure 14.23)

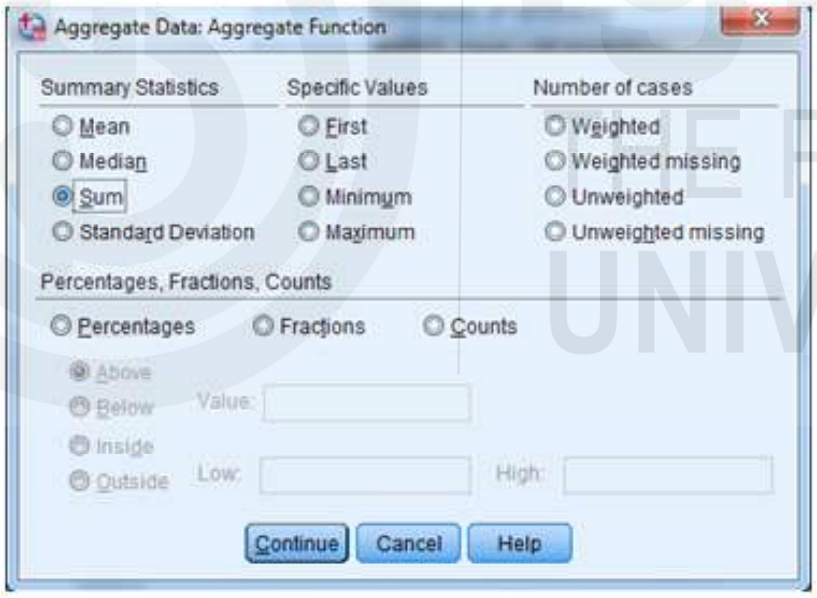

Fig.14.23: Aggregate Function dialog box

Select the function in the Aggregate Function dialog box. (For example, SUM)

- The selected function is assigned to the selected variables on the Aggregate Variables list.
- Click on Continue.
- If you want to change the variable name for the selected variable on the Aggregate Variables list and provide a descriptive variable label then Click Name and Label.
- Enter the new variable name.

- Select any one option from 'Save' category, Here in this example, 'create a new dataset containing only the aggregate variables' option is selected. For this, you can define a new dataset name for storing new aggregate data.
- Click OK button to finish aggregate data procedure.

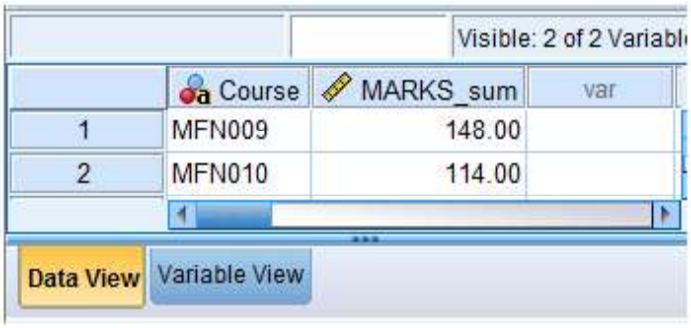

Fig.14.24: Aggregate data: New dataset

Now, new dataset will appear on your screen as shown in Figure 14.24.

14.10.3 Split File

When analyzing data, it is sometimes useful to temporarily "group" or "split" your data in order to compare results across different subsets. For this SPSS provides Split File option. Split File splits the data file into separate groups for analysis based on the values of one or more grouping variables. For example, if you select gender as the first grouping variable and minority as the second grouping variable, cases will be grouped by minority classification within each gender category.

A maximum of eight grouping variables can be defined. Cases need to be sorted out by grouping variables. Two options are available for comparative analysis. These are: compare groups and organize output by groups.

Compare groups

Split-file groups are presented together for comparison purposes. For pivot tables, a single pivot table is created and each split-file variable can be moved between table dimensions. For charts, a separate chart is created for each split-file group and the charts are displayed together in the Viewer.

Organize output by groups

All results from each procedure are displayed separately for each split-file group.

Fig.14.25: Data Menu

To split your dataset, you can follow these steps:

- Select Data option from menu bar and then select Split File. (see Figure 14.25)

- The Split File window (see Figure 14.26) will appear. By default, the dataset is not split according to any criteria; this is indicated by **Analyze all cases, do not create groups**.

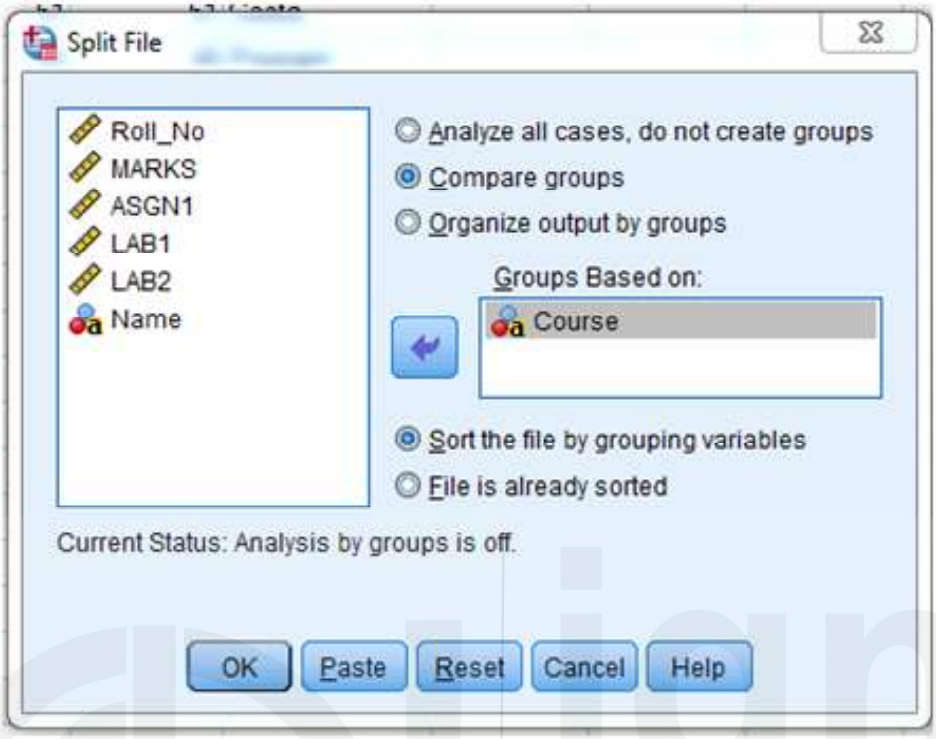

Fig.14.26: Split File window

- Select **Compare groups** or **Organize output by groups**.
- Select one or more grouping variables. Put the variable(s) under the '**Groups based on**' category (for e.g. Course)
- If the data file isn't already sorted by values of the grouping variables, select **Sort the file by grouping variables**.
- Click OK to complete split data procedure.
- After splitting the file, the only change you will see in the Data View is that data will be sorted in ascending order by the grouping variable(s) you selected.

14.10.4 Select Cases

Select case command can be used for selecting a random sub-sample or subgroup of cases based on specified criteria that includes variables and complex expressions.

The following criteria are used for Select Case command.

- Variable values and ranges
- Date and time ranges
- Case (row) numbers
- Arithmetic expressions
- Logical expressions
- Functions

The Select Cases is available under Data menu for selecting a subgroup.

Fig.14.27: Select Cases

Select offered several options which are given below: (see Figure 14.27)

- **All Cases** (default initial setting) Use it to remove a filter you have defined.
- **If condition is satisfied**: selects cases that satisfy a condition (logical expression) you specify after selecting the If button.
- **Random Sample of Cases:** This case selection is based on a random sample. The Sample button lets you specify e.g. the percentage of cases to select.
- **Based on time or case range**: It is mainly useful with time series data. Select a specific period or analysis.
- **Use filter variable:**Use the selected numeric variable from the data file as the filter variable.

Output section controls the treatment of unselected cases. You can choose one of the alternatives for the treatment of unselected cases such as Filter out unselected cases, Copy selected cases to a new dataset and Delete unselected cases.

Next, let us review the aspects linked with running a preliminary analysis.

14.11 RUNNING A PRELIMINARY ANALYSIS

Before running advanced statistical analysis, it is important that you understand the salient features of your data. Use of statistical applications on a data set, the behaviour of which is not known, can give misleading conclusions. The following section explains the six characteristics which must be examined for a given data set before attempting an advanced analysis.

14.11.1 Six Characteristics of a Dataset

One strong argument for using computers and graphical presentation of the data is the advantage of viewing the data in a variety of ways. Preliminary exploration

of data and its graphical presentation helps attain these objectives. The following characteristics will help you in deciding on the best plan for data management, analysis and presentation.

SPSS includes commands for analyzing of data along the following lines.

Shape: The shape of the data will be the main factor in determining what set of summary statistics best explains the data. Shape is commonly categorized as symmetric, left-skewed or right-skewed, and as uni-modal, bi-modal or multi-modal. Frequency distribution, plots and graphical presentation of data, histogram, P-P, Q-Q, scatter, Box-Plot are illustrative of the techniques that can be used for determining the shape of a data set. It is important that the user should have enough knowledge of the properties of various statistical distributions, their graphical presentations, characteristics and limitations.

Location: Location is simpler and more descriptive than measures of central tendency. Common measures of location are the mean and the median. Measures of central tendency also can be calculated for various sub-groups of a data set.

Spread: This measure describes the amount of variation in the data. Again approximatevalue is sufficient initially, with the measure of spread being informed by the shape of the data, and its intended use. Common measures of spread are variance, standard deviation and inter-quartile range. Percentile range is another measure which is usedfor measurement of dispersion.

Outliers: Outliers are data values that lie away from the general cluster of values. Each outlier needs to be examined to determine if it represents a possible value from the population being studied, in which case it should be retained, or if it is non-representative (or an error) in which case it should be excluded. You should properly weigh and carefully examine the behaviour of outliers before accepting or rejecting of an observation/case. The best choice to display when looking for outliers is Boxplot. Range, i.e., maximum and minimum values can also be used to examine the behaviour of outliers.

Clustering: Clustering implies that data tend to bunch around certain values. Clustering shows most clearly on a dot-plot. Histogram, stem and leaf analysis are also important procedures to examine the clustering pattern of a data set.

Association and relationship: Researchers often look for associative characteristics or similarities and dissimilarities in the behaviour of some variables. For example, achievement scores and hours of study may be positively correlated whereas the teacher motivation and drop-out rate may be negatively associated with each other. Correlation coefficient is the most commonly used measure for understanding the nature and magnitude of association between two variables.

You should be clear that association does not imply relationship. A relationship is defined by the cause and effect type of link. Normally, there is one dependent variable and one or more than one independent variable in the cause and effect relationship. Cause and effect relationship is captured through regression analysis.

The analysis of data along the above lines provides considerable insight into the nature of data and also helps researchers in understanding key relationships between variables. It is assumed that the relationships are of linear type. Non-

linear relationships can also be examined using non-linear techniques of analysis and also by using data transformation techniques as described next.

14.11.2 Data Transformation

Data transformation is a very useful aspect of SPSS. Using data transformation, you can collapse categories, recode the data and create new variables based on complex equations and conditional statements.

Data transformation (see Figure 14.28) provides some of the functions which are given below:

- Computing Variables
- Count Occurrences of Values within Cases
- Shift Values
- Random Number Generators
- Recode into Same Variables
- Recode into Different Variables
- Automatic Recode
- Rank Cases
- Date and Time Wizard
- Time Series Data Transformations
- Replace Missing Values

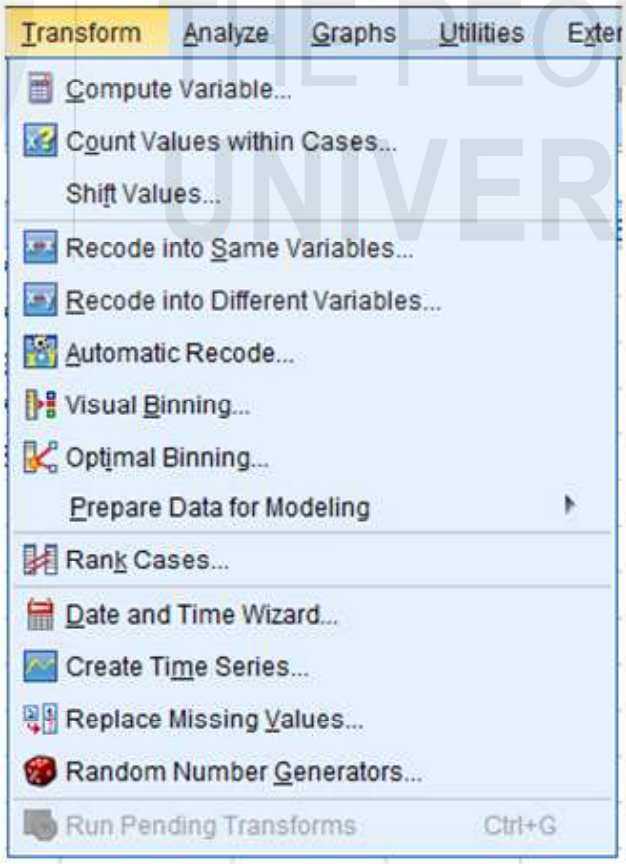

Fig.14.28 : Transform Menu

Compute variable

- Compute values for numeric or string variables
- Create new variables or replace the value of existing variables. For the new variables, you can specify the variable type and label.
- Compute values selectively for sub-sets of data based on logical conditions.
- Use built-in functions, statistical functions, distribution functions and string functions.

Rank Cases

Rank Cases command can be used to rank observations in an ascending or a descending order. Other options available for ranking cases are shown in the righthand panel of the Figure 14.29.

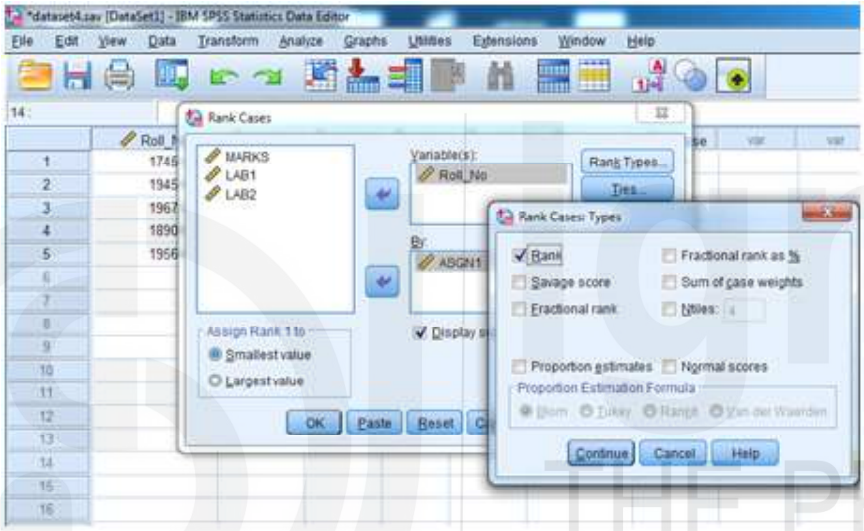

Fig. 14.29: Rank Cases

Count

Count is an important command available in SPSS and is used for counting occurrences of the same value(s) in a list if variables within the same case. For example, a survey might contain a list of books purchased (yes/no) by the students. You could count the number of 'yes' responses, or a new variable can be generated which gives the valueof count indicating the number of books bought.

Recoding (Transforming) Variables

Sometimes you will want to transform a variable by grouping its categories or values together. For example, you may want to change a continuous variable into a categorical variable, or you may want to merge the categories of a nominal variable. In SPSS, this type of transform is called recoding. In SPSS, there are some basic options for recoding variables such as **Recode into Different Variables** and **Recode into Same Variables**.

Next, we shall review the graphical presentation of data.

14.11.3 Graphical Presentation of Data

SPSS offers extensive facilities for viewing the data and its key features in high resolution charts and plots. SPSS has multiple ways of producing charts and

graphs. If you want to create charts and graphs in detailed manner, then you can choose Chart Builder option from Graphs menu.

For example: To produce a pie chart you can do the following:

- Click on the Graphs menu and select Pie option from submenu (see Figure 14.30)

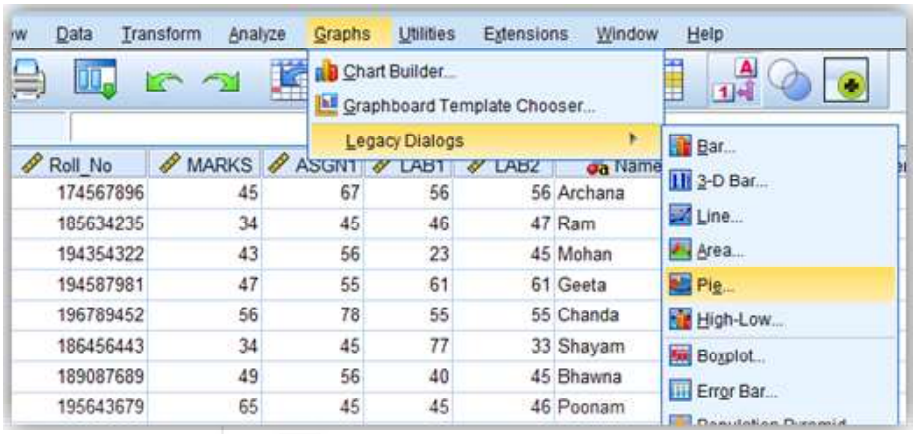

Fig. 14.30: Graphs menu

- This will open up the popup for Pie chart; here you can define data and click 'Define' button (see Figure 14.31)

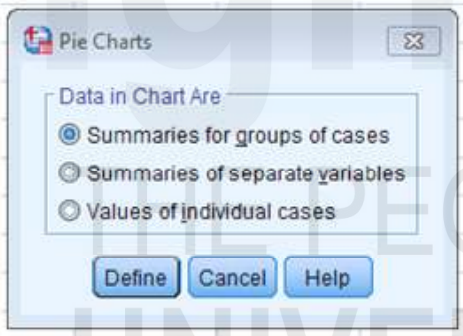

Fig. 14.31: Pie Chart menu (data option)

- This will open up the dialog box where you can define exactly what type of graph you would like.
- You can define Pie chart summaries for groups of cases (see Figure 14.32)

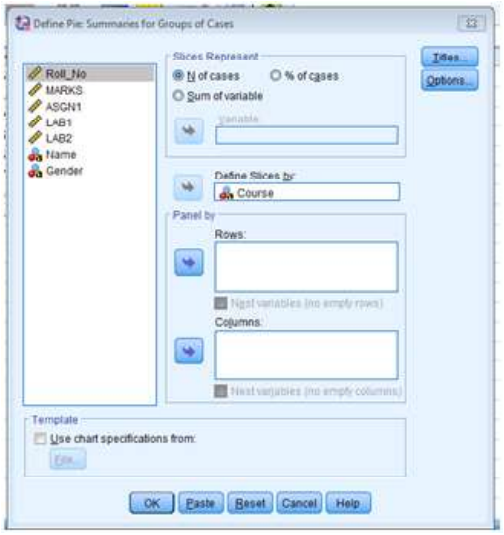

Fig. 14.32: Pie chart (summaries for groups of cases) window

- Click OK and get output in viewer window as shown below: (see Figure 14.33).

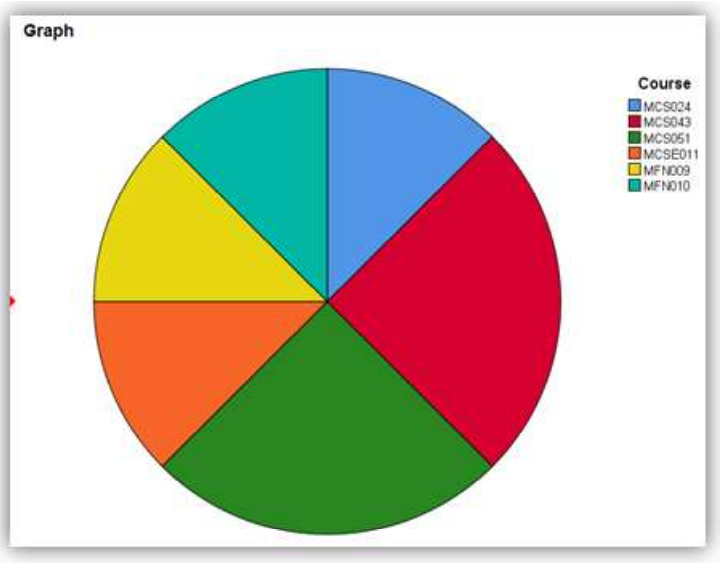

Fig. 14.33: Graph Output

14.12 UNDERSTANDING RELATIONSHIP BETWEEN VARIABLES: DATA ANALYSIS

The foregoing details focused on the techniques of analysis describing the behaviour of individual variables. However, most of the research studies require relationships between two or more variables to be examined. For example, one may be interested in questions like, "do the achievement scores of boys and girls in the same class differ?"

We shall review some of the various data analysis features specific to parametric and non-parametric tests.

14.12.1 Parametric Test

A parametric statistical test is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn. Under this sub-section we shall review frequency tables, cross-tabulations, correlations, simple regression. We begin with frequency distribution.

Here, we have discussed the procedure for frequency tables only. You are advised to consult the SPSS' users' manual and other statistical books for detailed discussion on the other tests.

We begin with frequency distribution.

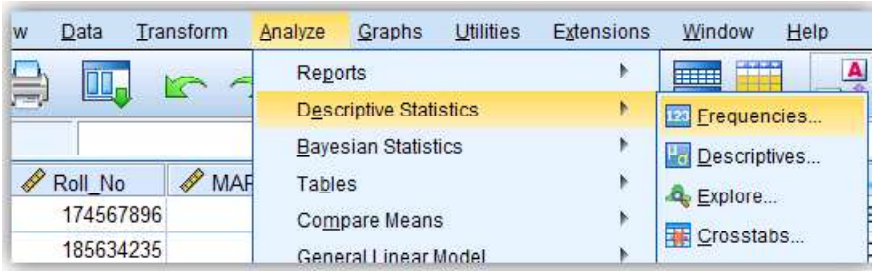

Fig.14.34: Analyze menu

To Analyze the data, choose Analyze From main menu and then Frequencies from Descriptive Statistics submenu(See Figure 14.34).

- Frequencies dialog box will display on screen. Select one or more variables (see Figure 14.35).

Fig. 14.35: Frequencies dialog box

- To do this click the variable "Gender" to select it for analysis, then optionally, you can:
- **Click Statistics** for descriptive statistics for quantitative variables.

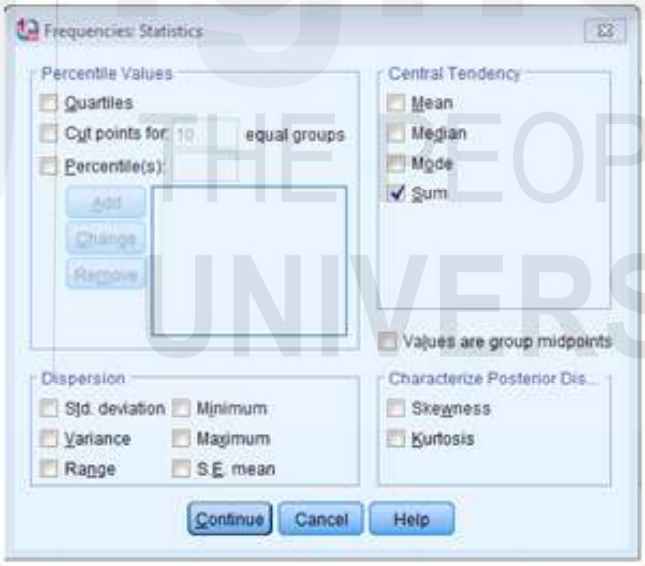

Fig. 14.36: Frequencies statistics dialog box

Click chart for bar chart, pie-charts and histogram.

- If you click statistics you will get a dialog box as shown in Figure 14.36.
- Now click the boxes for statistics you wish to apply for your data like, mean, standard deviation, etc. Then **click** the 'Continue' button.
- Now **click** the **OK** button, it automatically opens an Output – SPSS Viewer Window showing the Frequencies Tables as illustrated in Figure 14.37.

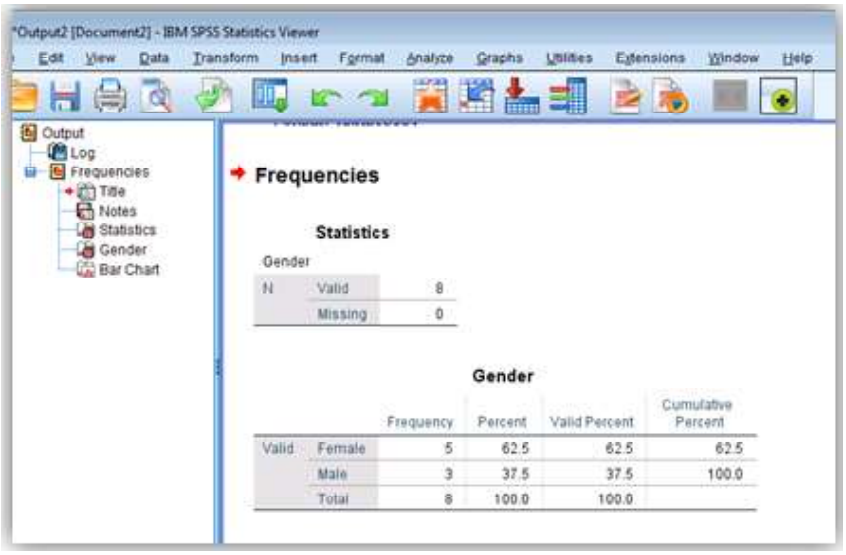

Fig.14.37: Output – SPSS Viewer Window

In case you wish to have charts then select Graphs from the menu in that you will see types of charts. Select the chart you wish to have and then the variable.

14.12.2 Non-Parametric Tests

The non-parametric test procedure provides several tests that do not require assumptions about the shape of the underlying distribution.

These include the following most commonly used tests:

- Chi-square test
- Binomial test
- Runs Test
- One sample Kolmogorov-Smirnov test
- Two independent Sample tests
- Tests for several independent samples
- Two related sample tests
- Tests for several related samples

Here, we shall discuss the procedure for Chi-square test only. You are advised to consult the SPSS' users' manual and other statistical books for detailed discussion on the other tests.

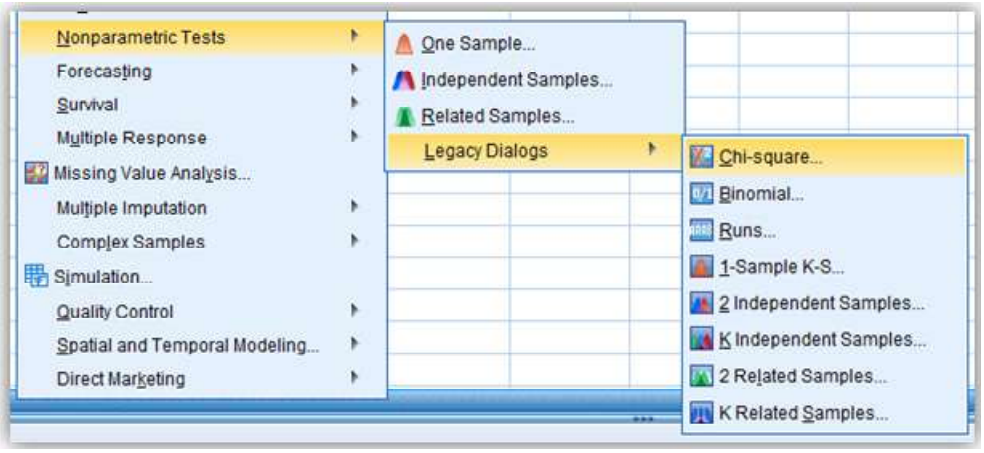

Fig. 14.38: Nonparametric Test

Chi-Square test

Chi-square test (refer to Figure 14.38) is the most commonly used test in social science research. The goodness of fit test compares the observed and the expected frequencies in each cell/category to test either that all categories contain the same proportion of values or that each category contains a user specified proportion of values.

Consider that a bag contains red, white and yellow balls. The chi-square test could be used to determine if a bag contains equal proportions of red, white and yellow balls.

You can do the following steps:

	Balls	Count
1	Red	1
2	Yellow	2
3	white	3
4	Yellow	2
5	Red	3
6	white	4
7	Red	5
8	Yellow	2
9	white	1
10	Red	3

Fig. 14.39: Sample Data

Create your data(refer to Figure 14.39)

- You can select **Nonparametric Tests** from **Analyze** main menu bar.
- Select **Chi-square** from the submenu (see figure 14.38)
- Chi-square Test dialog box will be open as shown in Figure 14.40.
- Put your variable in the test variable list (for e.g. Count)

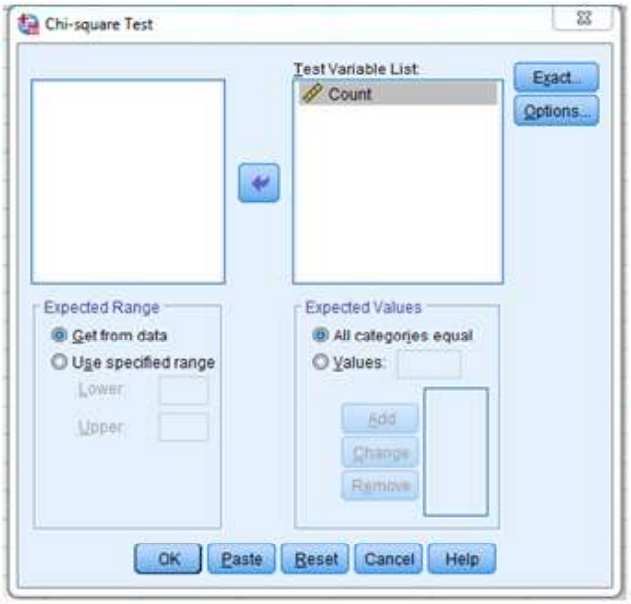

Fig. 14.40: Chi-square Test dialog box

- Click OK and get the output as shown below (see Figure 14.41)

Count

	Observed N	Expected N	Residual
1	2	2.0	.0
2	3	2.0	1.0
3	3	2.0	1.0
4	1	2.0	-1.0
5	1	2.0	-1.0
Total	10		

Test Statistics

	Count
Chi-Square	2.000[a]
df	4
Asymp. Sig.	.736

a. 5 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 2.0.

Fig. 14.41: Output -Chi-square Test

14.13 SPSS PRODUCTION FACILITY

The SPSS Production facility provides the ability to run SPSS in an automated mode. SPSS runs unattended and uninterrupted and terminates after executing the lastcommand. Production mode is useful if you run the same set of time-consuming analysis periodically.

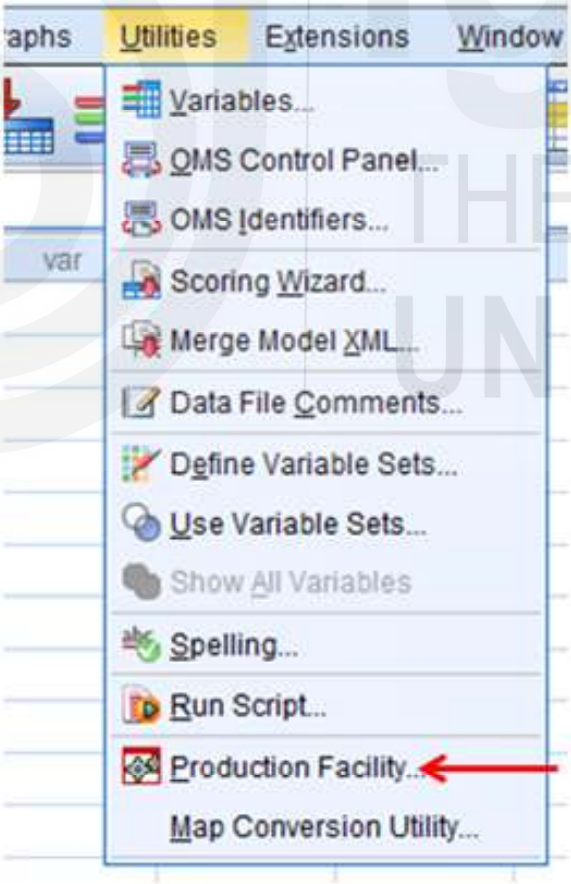

Fig. 14.42: SPSS Production facility

The SPSS Production facility uses command syntax file to tell SPSS about the commands to be executed. The command syntax file can be edited in a standard text editor.To run the SPSS Production facility, quit the SPSS if it is already running. SPSS

Production facility cannot be run when SPSS is running.

To create a production job, do the following:

- Select Production facilityoption from Utilities main menu bar.(see Figure 14.42)
- Production facility window will appear. (see Figure 14.43)
- Specify the syntax file that you want to use in the production job. Click Browse to select the Syntax File or create a new one.
- Next navigate to the Output section of the window. Here you need to choose where to save the SPSS output.
- Save the production file job.
- Run the production file job at any time.

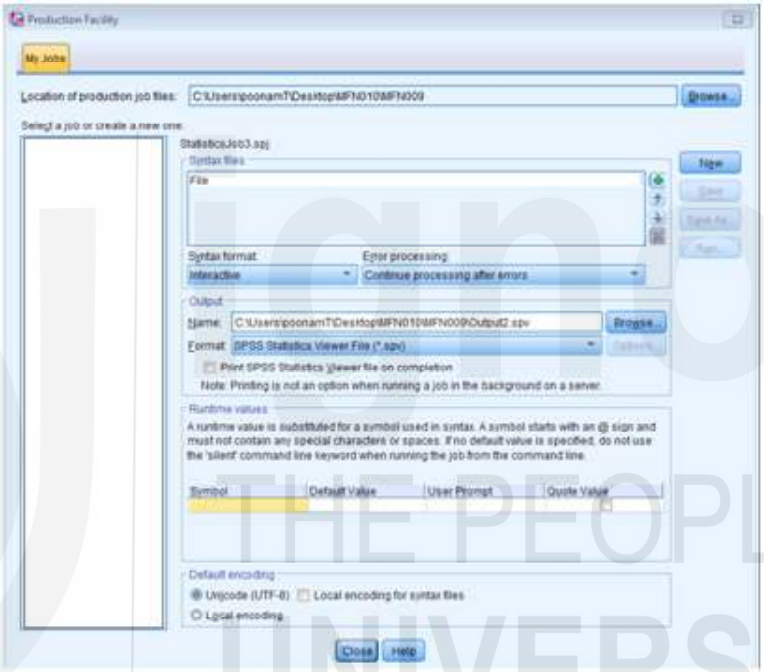

Fig. 14.43: Production facility window

Next, we shall review JMP Statistical Analysis System (SAS) and NUDIST package which are the other software packages available for analysis of quantitative and qualitative data, respectively.

14.14 JMP STATISTICAL ANALYSIS SYSTEM

JMP is a suite of computer programs for statistical analysis developed by the JMP business unit of SAS Institute. JMP helps you tackle your routine and difficult statistical problems. The software can be purchased in any of five configurations: JMP, JMP Pro, JMP Clinical, JMP Genomics and the JMP Graph Builder App.

Statistical Analysis capabilities in JMP® include:

- Univariate and multivariate linear and nonlinear analysis
- ANOVA, MANOVA and ANCOVA
- Time series analysis
- Multivariate optimization and Monte Carlo simulation
- Hierarchical and k-means clustering

14.15 NUDIST

NUDIST (**N**on-numerical **U**nstructured **D**ata **I**ndexing, **S**earching and **T**heorizing) is Computer programmes help in the analysis of qualitative data, especially in understanding a large (say 500 or more pages) text database. Studies using large databases such as ethnographies with extensive interviews, computer programmes provide an in valuable aid in research.

NUDIST programme was developed in Australia in 1991. This package is used for qualitative analysis of data. Here we present briefly the main features of this package. This software requires 4 megabytes of RAM and atleast 2 megabytes space for data files in your PC or MAC. In your PC it operates under windows (Cres well 1998).

As a researcher this software will help you to provide the following:

- **Storing and organizing files:** First establish document files and store information with the NUDIST programme. Document files consist of transcript from an interview, notes of observation or any article scanned from a newspaper.

- **Searching for themes:** Tag segments of text from all the documents that relate to a single idea or theme. For example, distance learners, in a study on effectiveness of distance education talk about the role of academic counsellors. The researcher can create a node in NUDIST as 'Role of Academic Counsellors'. Researcher will select text in the transcripts where learners have talked about this role and merge it into role of Academic Counsellors. Information can be retained in this node and researcher can take print in different ways in which learners talk about the role of academic counsellors.

- **Crossing themes:** Taking the same example of role of Counsellors, the researcher can relate this node to other nodes. Suppose the other node is qualifications of Counsellors. There are two categories like Graduate and Post Graduate. The researcher will ask NUDIST to cross the two categories, role of Counsellors and qualification of Counsellors to see for example whether there is any relation between graduate Counsellors and their role than the post graduate Counsellors and their role. NUDIST software generates information for a matrix with informationin the cells reflecting different perspectives.

- **Diagramming:** In this package; once the information is categorized, categories are identified. These categories are developed into nine visual picture of the categories that display their inter connectedness. This is called a tree diagram in NUDIST software. Tree diagram is a hierarchical tree of categories where root node is at the top and parents and siblings in the tree. This tree diagram is a useful device for discussing the data analysis of qualitative research in conferences.

- **Creating a template**: In a qualitative research, at the beginning of data analysis, the researcher will create a template which is a priori code book for organizing information.

14.16 LET US SUM UP

After using this unit, you are able to perform data entry and analysis and to create tables and graphs. In SPSS, three windows are commonly used viz., Data Editor Window, Syntax Editor Window and Output Viewer Window. The Data Editor window has Data View and Variable View. In Data View, you can see your actual data where as in variable view, you can define the Variable of the Data that you have. When you perform any command in SPSS, the Output Viewer window opens automatically and displays a log of the actions taken and the associated output.

SPSS Statistics can read and write data from Excel, plain text files and write to external relational database tables via ODBC and SQL. You can save statistical output in separate file format (*.spv file). The output can be exported to text or Microsoft Word, PDF, Excel, graphic image and other formats.

Using SPSS, you can merge data from two files in two different ways i.e. Add Cases and Add variables. You have also learned about the aggregate data. In statistics, aggregate data are data combined from several measurements. When data is aggregated, groups of observations are replaced with summary statistics based on those observations.

When analyzing data, it is sometimes useful to temporarily "group" or "split" your data in order to compare results across different subsets. For this, SPSS provides Split File option. Split File splits the data file into separate groups for analysis based on the values of one or more grouping variables.

Most of the research studies require relationships between two or more variables to be examined. For this some Parametric and non-Parametric tests are also discussed in this unit. For more test, you can go through the SPSS website.

You can use the SPSS Production facility to run SPSS in an automated mode.SPSS runs unattended and uninterrupted and terminates after executing the last command. This production mode is useful if you run the same set of time-consuming analysis periodically.

In this unit, some software is also introduced such as JMP Statistical Analysis System (SAS) software and NUDIST software for analysis of quantitative and qualitative data, respectively.

APPENDIX

Table I: Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distrance on the base line between mean and successive points laid off from the mean in units of standard deviation.

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	0.08	0.09
00	0000	00040	080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1054	1103	0141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	1611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3290	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4782	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4920	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	985	4985	4986	4987
3.0	4986.5	4986.9	4987.2	4987.8	4988.2	4992.4	4988.9	4989.3	4989.7	74990
3.1	4990.3	4990.6	4991	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	64992.9
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.992									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

Table II: Table of critical values of t

Degree of Freedom	Level of Significance			
	0.10	**0.05**	**0.02**	**0.01**
1	6.34	12.71	31.82	63.66
2	2.92	4.30	6.96	9.62
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.47	3.14	3.71
7	1.90	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.06
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.72	2.09	2.53	2.84
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.5	2.81
24	1.71	2.06	2.49	2.8
25	1.71	2.06	2.48	2.79
26	1.71	2.06	2.48	2.78
27	1.7	2.05	2.47	2.77
28	1.7	2.05	2.47	2.76
29	1.7	2.04	2.46	2.76
30	1.7	2.04	2.46	2.75
35	1.69	2.03	2.44	2.72
40	1.68	2.02	2.42	2.71
45	1.68	2.02	2.41	2.69
50	1.68	2.01	2.4	2.68
60	1.67	2.00	2.39	2.66
80	1.66	1.99	2.38	2.64
90	1.66	1.99	2.37	2.63
100	1.66	1.98	2.36	2.63
125	1.66	1.98	2.36	2.62
150	1.66	1.98	2.35	2.61
200	1.65	1.97	2.35	2.6
300	1.65	1.97	2.34	2.59
400	1.65	1.97	2.34	2.59
500	1.65	1.96	2.33	2.59
1000	1.65	1.96	2.33	2.58
∞	1.65	1.96	2.33	2.58

Table III: F-ratios for .05 (roman) and .01 **(bold face)** levels of significance

Degrees of freedom for smaller mean square	Degrees of freedom for greater mean square									
	1	2	3	4	5	6	8	12	24	∞
1	161.45 **4052.1**	199.5 **4999.93**	215.72 **5403.49**	224.57 **5625.14**	230.17 **5764.08**	233.97 **5859.39**	238.89 **5981.34**	243.91 **6105.83**	249.04 **6234.16**	254.32 **6366.48**
2	18.51 **98.49**	19 **99.01**	19.16 **99.17**	19.25 **99.25**	19.3 **99.33**	19.33 **99.33**	19.37 **99.36**	19.41 **99.42**	19.45 **99.46**	19.50 **99.50**
3	10.13 **34.12**	9.55 **30.81**	9.28 **29.46**	9.12 **28.71**	9.01 **28.24**	8.94 **27.91**	8.84 **27.49**	8.74 **27.05**	8.64 **26.6**	8.53 **26.12**
4	7.71 **21.20**	6.94 **18**	6.59 **16.69**	6.39 **15.98**	6.26 **15.52**	6.16 **15.21**	6.04 **14.8**	5.91 **14.37**	5.77 **13.93**	5.63 **13.46**
5	6.61 **16.26**	5.79 **13.27**	5.41 **12.06**	5.19 **11.39**	5.05 **10.97**	4.95 **10.67**	4.82 **10.27**	4.68 **9.89**	4.53 **9.47**	4.36 **9.02**
6	5.99 **13.74**	5.14 **10.92**	4.76 **9.78**	4.53 **9.15**	4.39 **8.75**	4.28 **8.47**	4.15 **8.1**	4 **7.72**	3.84 **7.31**	3.67 **6.88**
7	5.59 **12.25**	4.74 **9.55**	4.35 **8.45**	4.12 **7.85**	3.97 **7.64**	3.87 **7.19**	3.73 **6.84**	3.57 **6.47**	3.41 **6.07**	3.23 **5.65**
8	5.32 **11.26**	4.46 **8.65**	4.07 **7.59**	3.84 **7.01**	3.69 **6.63**	3.58 **6.37**	3.41 **6.03**	3.28 **5.67**	3.12 **5.28**	2.93 **4.86**
9	5.12 **10.56**	4.26 **8.02**	3.86 **6.99**	3.64 **6.42**	3.48 **6.06**	2.37 **8.8**	3.23 **5.47**	3.07 **5.11**	2.9 **4.73**	2.71 **4.31**
10	4.96 **10.04**	4.1 **7.56**	3.71 **6.55**	3.48 **5.99**	3.33 **5.64**	3.22 **5.39**	3.07 **5.06**	2.91 **4.71**	2.74 **4.33**	2.54 **3.91**
11	4.84 **9.65**	3.98 **7.2**	3.59 **6.22**	3.36 **5.67**	3.2 **5.32**	3.09 **5.07**	2.95 **4.74**	2.79 **4.4**	2.61 **4.02**	2.4 **3.6**

Table III: Continued

	Degrees of freedom for greater mean square									
	1	2	3	4	5	6	8	12	24	∞
12	4.75	3.88	3.49	3.26	3.11	3	2.85	2.69	2.5	2.30
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.5**	**4.16**	**3.78**	**3.36**
13	4.67	3.8	3.41	3.18	3.02	2.92	2.77	2.6	2.42	2.21
	9.07	**6.7**	**5.74**	**5.2**	**4.86**	**4.62**	**4.3**	**3.96**	**3.59**	**3.16**
14	4.60	3.74	3.34	3.11	2.96	2.85	2.7	2.53	2.35	2.13
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.14**	**3.8**	**3.43**	**3.00**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.00**	**3.67**	**3.29**	**2.87**
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.2**	**3.89**	**3.55**	**3.18**	**2.75**
17	4.45	3.59	3.20	2.96	2.81	2.7	2.55	2.38	2.19	1.96
	8.4	**6.11**	**5.18**	**4.67**	**4.34**	**4.1**	**3.79**	**3.45**	**3.08**	**2.65**
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.71**	**3.37**	**3.01**	**2.57**
19	4.38	3.52	3.13	2.9	2.74	2.63	2.48	2.31	2.11	1.88
	8.18	**5.93**	**5.01**	**4.5**	**4.17**	**3.94**	**3.63**	**3.3**	**2.92**	**2.49**
20	4.35	3.49	3.10	2.87	2.71	2.6	2.45	2.28	2.08	1.84
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.56**	**3.23**	**2.86**	**2.42**
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.51**	**3.17**	**2.80**	**2.36**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.4	2.23	2.03	1.78
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.75**	**3.45**	**3.12**	**2.75**	**2.30**
23	4.28	3.42	3.03	2.8	2.64	2.53	2.38	2.2	2.00	1.76
	7.88	**5.66**	**4.76**	**4.46**	**3.94**	**3.71**	**3.41**	**3.07**	**2.70**	**2.26**

Table III: Continued

	Degrees of freedom for greater mean square									
	1	2	3	4	5	6	8	12	24	∞
24	4026	3.4	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
	7.82	**5.61**	**4.72**	**4.22**	**3.9**	**3.67**	**3.36**	**3.03**	**2.66**	**4.21**
25	4.24	3.38	2.99	2.76	2.6	2.49	2.34	2.16	1.96	1.71
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.32**	**2.99**	**2.62**	**2.17**
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.29**	**2.96**	**2.58**	**2.13**
27	4.21	3.35	2.96	2.73	2.57	2.46	2.3	2.13	1.93	1.67
	7.68	**5.49**	**4.6**	**4.11**	**3.78**	**3.56**	**3.26**	**2.93**	**2.55**	**2.1**
28	4.2	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
	7.64	**5.45**	**4.57**	**4.07**	**3.75**	**3.53**	**3.23**	**2.9**	**2.52**	**2.06**
29	4.18	3.33	2.93	2.7	2.54	2.43	2.28	2.1	1.9	1.64
	7.6	**5.42**	**4.54**	**4.04**	**3.73**	**3.5**	**3.2**	**2.87**	**2.49**	**2.08**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
	7.56	**5.39**	**4.51**	**4.02**	**3.7**	**3.47**	**3.17**	**2.84**	**2.47**	**2.01**
35	4.12	3.26	2.87	2.64	2.48	2.37	2.22	2.04	1.83	1.57
	7.42	**5.27**	**4.4**	**3.91**	**3.59**	**3.37**	**3.07**	**2.74**	**2.37**	**1.9**
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2	1.79	1.52
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**2.99**	**2.66**	**2.29**	**1.82**
45	4.06	3.21	2.81	2.58	2.42	2.31	2.15	1.97	1.76	1.48
	7.23	**5.11**	**4.25**	**3.77**	**3.45**	**3.23**	**2.94**	**2.61**	**2.23**	**1.75**
50	4.03	3.18	2.79	2.56	2.4	2.29	2.13	1.95	1.74	1.44
	7.17	**5.06**	**4.2**	**3.72**	**3.41**	**3.19**	**2.89**	**2.56**	**2.18**	**1.68**
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.7	1.39
	7.08	**4.98**	**3.65**	**3.65**	**3.34**	**3.12**	**2.82**	**2.5**	**2.12**	**1.60**

Table III: Continued

	Degrees of freedom for greater mean square									
	1	2	3	4	5	6	8	12	24	∞
70	3.98	3.13	2.74	2.5	2.35	2.23	2.07	1.89	1.67	1.35
	7.01	**4.92**	**4.07**	**3.6**	**3.29**	**3.07**	**2.78**	**2.45**	**2.07**	**1.53**
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.88	1.65	1.31
	6.96	**4.88**	**4.04**	**3.56**	**3.26**	**3.04**	**2.74**	**2.42**	**2.03**	**1.47**
90	3.95	3.1	2.71	2.47	2.32	2.2	2.04	1.86	1.64	1.28
	6.92	**4.85**	**4.01**	**3.53**	**3.23**	**3.01**	**2.72**	**2.39**	**2**	**1.43**
100	3.94	3.09	2.7	2.46	2.3	2.19	2.03	1.85	1.63	1.26
	6.90	**4.82**	**3.98**	**3.51**	**3.21**	**2.99**	**2.69**	**2.37**	**1.98**	**1.39**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.01	1.83	1.6	1.21
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.66**	**2.33**	**1.94**	**1.32**
150	3.90	3.06	2.66	2.43	2.27	2.16	2	1.82	1.5	1.18
	6.81	**4.75**	**3.91**	**3.45**	**3.14**	**2.92**	**2.63**	**2.31**	**1.92**	**1.27**
200	3.89	3.04	2.65	2.42	2.26	2.14	1.98	1.8	1.57	1.14
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.89**	**2.6**	**2.28**	**1.88**	**1.21**
300	3.87	3.03	2.64	2.41	2.25	2.13	1.97	1.79	1.55	1.10
	6.72	**4.68**	**3.85**	**3.38**	**3.08**	**2.86**	**2.57**	**2.24**	**1.85**	**1.14**
400	3.86	3.02	2.63	2.4	2.24	2.12	1.96	1.78	1.54	1.07
	6.70	**4.66**	**3.83**	**3.37**	**3.06**	**2.85**	**2.56**	**2.23**	**1.84**	**1.11**
500	3.86	3.01	2.62	2.39	2.23	2.11	1.96	1.77	1.54	1.06
	6.69	**4.065**	**3.82**	**3.36**	**3.05**	**2.84**	**2.55**	**2.2**	**1.83**	**1.08**
1000	3.85	3.00	2.61	2.38	2.22	2.1	1.95	2.76	1.53	1.03
	6.66	**4.63**	**3.80**	**3.34**	**3.04**	**2.82**	**2.53**	**2.2**	**1.81**	**1.04**
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.51**	**2.18**	**1.79**	**1.00**

Table IV: χ^2 Table. P give the probability of exceeding the tabulated value of χ^2 for the specified number of degrees of freedom (df). The values of χ^2 are printed in the body of the table.

df	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	.00393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	3.412	6.635
2	0.103	0.211	0.446	0.713	1.383	2.408	3.219	4.605	5.991	7.824	9.21
3	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	1.145	1.61	2.343	3	4.351	0.064	7.289	9.236	11.07	13.388	15.086
6	1.635	2.204	3.07	3.828	5.348	7.231	8.558	10.645	12.592	15.083	16.812
7	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	2.733	3.49	4.594	5.527	7.344	9.524	11.03	13.362	15.507	18.168	20.09
9	3.325	4.168	5.38	6.393	8.313	10.656	12.242	14.684	16.919	19.679	21.666
10	3.94	4.865	6.179	7.267	9.312	11.781	13.442	15.987	18.307	21.161	23.209
11	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	5.226	6.304	7.807	9.034	11.31	14.011	15.812	18.549	21.026	24.054	26.217
13	5.892	7.042	8.634	9.926	12.34	15.119	16.985	19.812	22.362	25.472	37.688
14	6.571	7.79	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.241
15	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	2.633	32
17	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.76	27.587	30.995	33.409
18	9.39	10.865	12.857	14.44	17.338	20.601	22.76	25.989	28.869	32.346	34.805
19	10.117	11.651	13.716	15.352	18.338	21.689	23.9	27.204	30.144	33.687	36.191
20	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.41	35.02	37.566
21	11.591	13.24	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	13.091	14.848	17.187	19.021	22.337	26.015	28.429	32.007	35.172	38.968	41.638
24	13.848	15.66	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.27	42.98
25	14.611	16.473	18.94	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	15.379	17.292	19.82	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.14	46.963
28	16.928	18.94	21.588	23.647	27.336	31.391	34.027	37.913	41.337	45.419	48.278
29	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	18.493	20.599	23.364	25.508	29.336	33.53	36.25	40.256	43.773	47.962	50.892

Table 4: Random Digits[α]

11339	19233	50911	14209	39594	68368	97742	36252	27671	55091
96971	19968	31709	40197	16313	80020	01588	21654	50328	04577
07779	47712	33846	84716	49870	59670	46946	71716	50623	38681
71675	95993	08790	13241	71260	16558	83316	68482	10294	45137
12804	72742	16237	72550	10570	31470	92612	94917	48822	79794
14835	56263	53062	71543	67632	30337	28739	17582	40924	32434
15544	14327	07580	48813	30161	10746	96470	60680	63507	14435
92230	41243	90765	08867	08038	05038	10908	00633	21740	55450
33564	93563	10770	10595	71323	84243	09402	62877	49762	56151
84461	55618	40570	72906	30794	491444	65239	21788	38288	29180
91645	42451	83776	99246	45548	02457	74804	49536	89815	74285
78305	63797	26995	23146	56071	97081	22376	09819	56855	97424
97888	55122	65545	02904	40042	70653	24483	31258	96475	77668
67286	09001	09718	67231	54033	24185	52097	78713	95910	84400
53610	59459	894945	72102	66595	02198	26968	88467	46939	52318
52965	76189	68892	64541	02225	09603	59304	38179	75920	80486
25336	39735	25594	50557	96257	59700	27715	42432	27652	88151
73078	44371	77616	49296	55882	71507	30168	31876	28283	53424
31797	52244	38354	47800	48454	43304	14256	74281	82279	28882
47772	22798	36910	39986	34033	39868	24009	97123	59151	27583
54153	70832	37575	31898	39212	63993	05419	77565	73150	98537
93745	99871	37129	55032	94444	17884	27082	23502	06136	89476
81676	51330	58828	74199	87214	13727	80539	95037	73536	16862
79788	02193	33250	05865	53018	62394	56997	41534	01953	13763
92112	61235	68760	61201	02189	09424	24156	10368	26527	89107
87542	28171	45150	75523	66790	63963	13903	68498	02981	25219
37535	48342	48943	07719	20407	33748	93650	39356	01011	22099
95957	96668	69380	49091	90182	13205	71802	35482	27973	46814
34642	85350	53361	63940	79546	89956	96836	81313	80712	73572
50413	31008	09231	46516	61672	79954	01291	72278	55658	84893
53312	73768	59931	55182	43761	59424	79775	17772	41552	45236
16302	64092	76045	28958	21182	30050	96256	85737	86962	27067
96357	98654	01909	58799	87374	53184	87233	55275	59572	56476
38529	89095	89538	15600	33687	86353	61917	63876	52367	79032
45939	05014	06099	76041	57638	55342	41269	96173	94872	35605
02300	23739	68485	98567	77035	91533	62500	31548	09511	80252
59750	14131	24973	05962	83215	25950	43867	75213	21500	17758
21285	53607	82657	22053	29996	04729	48917	72091	57336	18476
93703	60164	19090	63030	88931	84439	94747	77982	61932	21928
15576	76654	19775	77518	43259	82790	08193	63007	68824	75315
12752	33321	69796	03625	37328	75200	77262	99004	96705	15540
89038	53455	93322	25069	88186	45026	31020	52540	10838	72490
62411	56968	08379	40159	27419	12024	99694	68668	73039	87682
45853	68103	38927	77105	65241	70387	01634	59665	30512	66161
84558	24272	84355	00116	68344	92805	52618	51584	75964	53021
45272	58388	69131	61075	80192	45959	76992	19210	27126	45525
68015	99001	11832	39832	80462	70468	89929	55695	777524	20675
13263	92240	89559	66545	06433	38634	36645	22350	81169	97417
66309	31466	97705	46996	69059	33771	95004	89037	38054	80853
56348	05291	38713	82303	26293	61319	45285	75784	50043	44438

[α]From pp.99 and 100 of The Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates*, The Free Press, Glencoe, Ill., 1955, with permission of the Rand Corporation.

93108	77033	68325	10160	38667	62441	87023	94372	06164	30700
28271	08589	83279	48838	60935	70541	53814	95588	05832	80235
21841	35545	11148	34775	17308	88034	97765	35959	52843	44895
22025	79554	19698	25255	50283	94037	57463	92925	12042	91414
09210	20779	02994	02258	86978	85092	54052	18354	20914	28460
90552	71129	03621	20517	16908	086668	29916	51537	93658	29525
01130	06995	20258	10351	99248	51660	38861	49668	74742	47181
22604	56719	21784	68788	38358	59827	19270	99287	81193	43366
06690	01800	34272	65497	94891	14537	91358	21587	95765	72605
59809	69982	71809	64984	48709	43991	24987	69246	86400	29559
56475	02726	58511	95405	70293	84971	06676	44075	32338	31980
02730	34870	83209	03138	07715	31557	55242	61308	26507	06186
74482	33990	13509	92588	10462	76546	46097	01825	20153	36271
19793	22487	94238	81054	95488	23617	15539	94335	73822	93481
19020	27856	60526	24144	98021	60564	46373	86928	52135	47919
69565	60635	65709	77887	42766	86698	14004	94577	27936	47220
69274	23208	61035	84263	15034	28717	76146	22021	23779	98562
83658	14204	09445	41081	49630	34215	89806	40930	97194	21747
78612	51102	66826	40430	54072	62164	68977	95583	11765	81072
14980	74158	78216	38985	60838	82836	42777	85321	90463	11813
63172	28010	29405	91554	75195	51183	65805	87525	35952	83204
71167	37984	52737	06869	38122	953222	41356	19391	96787	64410
78530	56410	19195	34434	83712	50397	80920	15464	81350	18673
98324	03774	07573	67864	06497	20758	83454	22756	83959	96347
55793	30055	08373	32652	02654	75980	02095	87545	88815	80086
05674	34471	61967	91266	38814	44728	32455	17057	08339	93997
15643	22245	07592	22078	73628	60902	41561	54608	41023	98345
66750	19609	70358	03622	64898	82220	69304	46235	97332	64539
42320	74314	50222	82339	51564	42885	50482	98501	02245	88990
73752	73818	15470	04914	24936	65514	56633	72030	30856	85183
97546	02188	46373	21486	28221	08155	23486	66134	88799	49496
32569	52162	38444	42004	78011	16909	94194	79732	47114	23919
36048	93973	82596	28739	86985	58144	65007	08786	14826	04896
40455	36702	38965	56042	80023	28169	04174	65533	52718	55255
33597	47071	55618	51796	71027	46690	08002	45066	02870	60012
22828	96380	35883	15910	17211	42358	14056	55438	98148	35384
00631	95925	19324	31497	88118	06283	84596	72091	53987	01477
75722	36478	07634	63114	27164	15467	03983	09141	60562	65725
80577	01771	61510	17099	28731	41426	18853	41523	14914	76661
10524	20900	65463	83680	05005	11611	64426	59065	06758	02892
93815	69446	75253	51915	97839	75427	90685	60352	96288	34248
81867	97119	93446	20862	46591	97677	42704	13718	44975	67145
64649	07689	16711	12169	15238	74106	60655	56289	74166	78561
55768	09210	52439	33355	57884	36791	00853	49969	74814	09270
38080	49460	48137	61589	42742	92035	21766	19435	92579	27683
22360	16332	05343	34613	24013	98831	17157	44089	07366	66196
40521	09057	00239	51284	71556	22605	41293	54854	39736	05113
19292	69862	59951	49644	53486	28244	20714	56030	39292	45166
79504	40078	06838	05509	68581	39400	85615	52314	83202	40313
64138	27983	84048	42631	58658	62243	82572	45211	37060	15017