UNIT 14 DATA MANAGEMENT, ANALYSIS AND PRESENTATION

Structure

14.1 INTRODUCTION

Decision making which is based on information has now come to stay.

Statistical analysis and interpretation of data now forms an integral part of policy making, strategic planning and related support services in the area of health and nutrition. 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 desktops, which are often called PC. Earlier, the software's which could only be run on large mainframe computers can now be run with considerable ease on the PCs. The use of MS Excel software package that 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).

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 present unit will help you get acquainted with the statistical capabilities of SPSS for Windows. We will also introduce in this unit the statistical analysis system (SAS) software for quantitative data analysis. For qualitative data analysis we will introduce software called NUDIST.

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.

14.2 INTRODUCTION TO SPSS

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 (Statistical Package for Social Sciences) 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 analyze 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.

14.3 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.

SPSS Professional Statistics provides techniques to examine similarities and dissimilarities in data and to classify data, identify underlying dimensions in a data set. It includes procedures for cluster, k-cluster, discriminate, factor, multi- dimensional scaling, and proximity and reliability analysis.

SPSS Advanced Statistics includes procedures for logistic regression, log-linear analysis, multivariate analysis and analysis of variance. This module also includes procedures for constrained non-linear regression, probit, Cox and actuarial survival analysis.

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

SPSS Trends performs comprehensive forecasting and time series analysis with multiple curve fitting models, smoothing models and methods for estimation of autoregressive functions.

SPSS categories performs conjoint analysis and optimal scaling procedures, including correspondence analysis.

SPSS Chaid provides simplified tabular analysis of categories data, develops predictive models, screens out extraneous predictor variables, and produces easy-to-read tree diagrams that segment a population into sub-groups that share similar characteristics.

Recently, the SPSS corporation announced the release of SPSS version 15.0. Many new add-on products have also been launched in the recent months. 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.

The present unit discusses some of the commonly used data, management techniques and statistical procedures using SPSS 11.5 version since new features are added almost daily, you are advised to check for these details on the currently installed version 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. Here in this unit we will attempt to provide you with procedures that are most commonly used with SPSS Release 11.5. In case these are not available on your version of SPSS, please consult the relevant SPSS authorized representative or the WWW site of the SPSS corporation.

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

14.4 GET YOURSELF ACQUAINTED WITH SPSS

The SPSS for Windows can be run from Windows 3.x or Window 95 through Windows 98 or later operating systems. Unix, Mac and mainframe versions of the SPSS software are also available. The illustrations in this unit are based on SPSS version for Window 95/98/NT operating systems.

Starting SPSS

The SPSS for Windows uses graphical environment, descriptive menus and simple dialog boxes to do most of the work. It produces three type of files, namely *data files, chart files* and *text files.*

To start SPSS, click the *start* button on your computer. On the start menu that appears, click *Program*. Another menu appears on the right of the *start* menu. If there is an entry marked SPSS, that's the one you want to click. If there isn't, click the program group where SPSS was installed and an entry marked SPSS will appear. Click the SPSS 11.5 (or which ever version entry). You will know when the SPSS has started and an SPSS Data Editor window appears. To begin with, the SPSS data editor window will be empty and a number of menus will appear on the top of the window. We can start the operations by loading a data set or by creating a new file for which data is to be entered from the data editor window. The data can also be imported from other programs like Dbase, ASCII, Excel and Lotus, we will learn about this in a little while from now.

Exiting SPSS

Make sure that all SPSS and other files are saved before quitting the program. You should 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 spreadsheets 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. It is suggested that ink jet or laser jet printers should be used for printing graphs and charts. Tabular data can be easily printed using a Dot matrix Printer.

Operating Windows in SPSS

There are seven type of Windows in SPSS which are frequently referred to during the data management and analysis stages. These are:

Data Editor

As mentioned earlier, the data editor window opens automatically as soon the SPSS gets loaded. To begin with, the data editor does not contain any data. The file containing the data for analysis has to be loaded with the help of 'file' menu sub-commands by using various options available for this purpose. The contents of the active data file are displayed in the data editor window. Only one data editor window will be active at a time. No statistical operations can be performed until some data is loaded into data editor.

Output Navigator

All SPSS messages, statistical results, tables and charts are displayed in the output navigator. The output navigator can be opened/closed using the File Open/New Command. The output in the navigator window can be edited and saved for future reference. The Output Navigator opens automatically, the first time some output is generated. The user can customize the presentation of reports and tables displayed in the Output Navigator. The output can be directly imported into reports prepared under word processing packages, and the output files are saved with an extension xxxx.spo.

Pivot Tables

The output shown in the Output Navigator can be modified in many ways using the Edit and Pivot Table Option, which can be used to edit text, swap rows and columns, add colour, prepare custom made reports/output, create and display selectively multi-dimensional tables. The results can be selectively hidden and shown using features available in Pivot Tables.

Graphics

The Chart Editor helps in switching between various type of charts, in swapping of X - Y axis, changing colour and providing facilities for presenting data and results through various type of graphical presentations. It is useful for customizing the charts to highlight specific features of the charts and maps.

Text Editor

The text output not displayed in the Pivot Tables can be modified with the help of Text Editor. It works like an ordinary Text Editor. The output can be saved for future reference or sharing purposes.

Syntax Editor

The Syntax Editor can be opened and closed like any other file using the File Open/ New command. The use of Syntax File is recommended when the same type of analysis is to performed at frequent intervals of time or on a large number of data files. Using Syntax File for such purposes automates complex analysis and also avoids errors due to frequent typing of the same command. The commands can be pasted on the Syntax files using a particular command and pastes buttons from the menu. Experienced users can directly type the commands in the Syntax window. To run the syntax, select the commands to be executed and click on the run button at the top of the syntax window. All or some selected commands from the Syntax File will be executed. The Syntax File is saved as xx.sps.

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.

14.5 MENU COMMANDS AND SUB-COMMANDS

Most of the commands can be executed by making appropriate selections from the menu bar. Some additional commands and procedures are available only through the Syntax Window. The SPSS user manuals provide a comprehensive list of commands, which are not available through menu driven options. If you want a comprehensive overview of the basics of SPSS, there is an on-line tutorial, as extensive help on SPSS is available by using the 'Help' menu command. The CD version of the software contains an additional demo module.

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. We will present below in Table 14.1. The menu and sub-menus of the Data Editor window. Refer to Figure 14.1 which highlights the SPSS data editor.

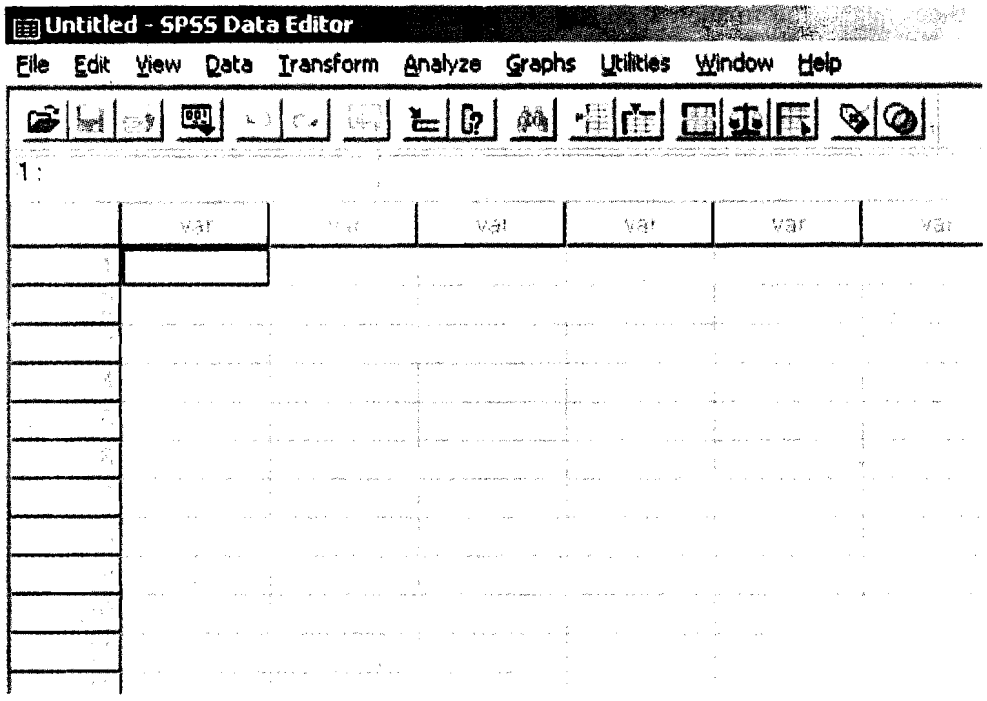

Figure 14.1: SPSS data editor shows the data editor menus. Each command in the main menu has a number of sub-commands.

Table 14.1 : Components of data editor menu

Menu	Function/sub-commands
File	Open and Save data file, to import data created in other formats like Lotus, Excel, Dbf etc. Print control functions like page setup, printer setup and associated functions. ASCII data can also be read into SPSS.
Edit	These functions are similar to those available in general packages. These include undo, redo, cut, copy, paste, paste variable, find, find and replace. Option setting for the SPSS are controlled through Edit menu.
View	Customize tool bars, Fonts, grid and display of data, displays option for showing value labels.
Data	This is a very important menu as far as management of the data is concerned. Variable definition, inserting new variables, transposing templates, aggregating and merging of data files, splitting data files for 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, ANOVA, Log linear regression, discriminate analysis, factor analysis, non-parametric tests and time series analysis are undertaken using analyze menu.

Graphs	Includes options for generating various type of custom made graphics like bar, pie, area, X- Y and high-low charts, praeto, control charts, box-plots, histograms, P-P and Q-Q charts and time series representation of data.
Utilities	Information about variables, information on working a data file, run scripts and define sets are some of the important functions carried out through Utilities command.
Window	Windows menu are used to switch between SPSS windows.
Help	Context specific help through dialog boxes, demo of the software, and information about the software are some of the important options under Help command. It provides a connection for the SPSS home page. The statistical coach included in the help module is very useful in understanding various stages of executing a procedure.

Setting The Options

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 14.2. Make the appropriate changes to set the options according to your choice.

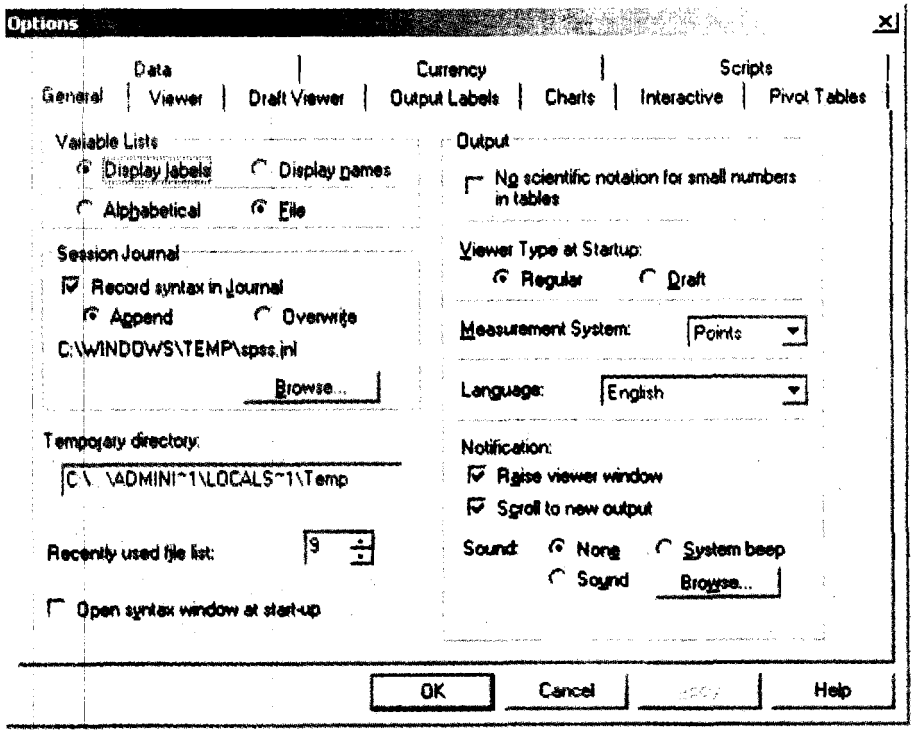

Figure 14.2: SPSS options

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

14.6 BASIC STEPS IN DATA ANALYSIS

There are five basic steps involved in data analysis using SPSS. These are shown in the Figure 14.3.

Let us review these steps.

Bring your data into SPSS: You can bring your data into SPSS in the following ways:

- Enter data directly into SPSS Data Editor
- Open previously saved SPSS data file

- Read a spreadsheet data into SPSS data editor
- Import data from DBF files
- Import data-from RDBMS packages like Access, Oracle, Power Builder, etc.

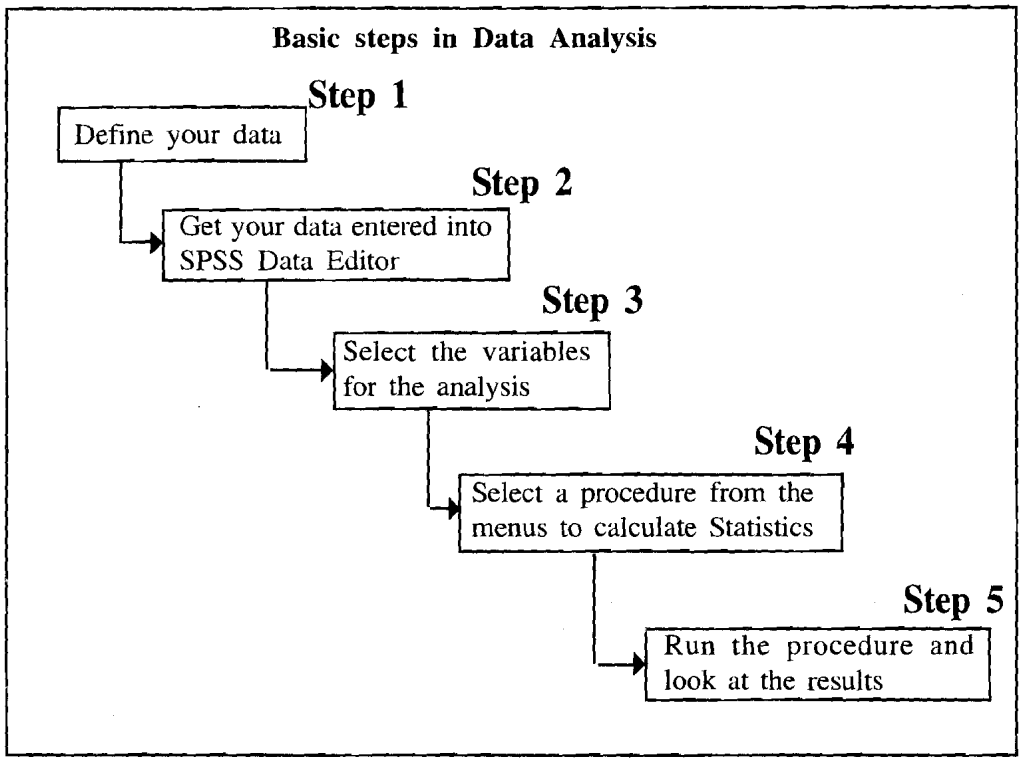

Figure 14.3: Steps in data analysis

Select the Variables: The variables in the active file are listed each time a dialog box is opened. Select the appropriate variables for the selected procedure. Selection of at least one variable is necessary to run a statistical procedure. The variables may be numeric, string, date or logical. You should be aware that string variables cannot be manipulated to the same extent as the numeric variables.

Select a Procedure from Menus: Before embarking on a statistical analysis, it is advised that you are clear as to what analysis is to be performed. Select the corresponding procedure to work on the data or create charts or tables using the selected procedure.

The command could either be directly executed or pasted on a Syntax Window. As mentioned earlier, pasting the command on the Syntax Window will be useful for undertaking batch processing or for subsequent use, especially where the same type of repetitive analysis required. Pasting the command will not lead to its execution. The command has to be selected and executed using the run command.

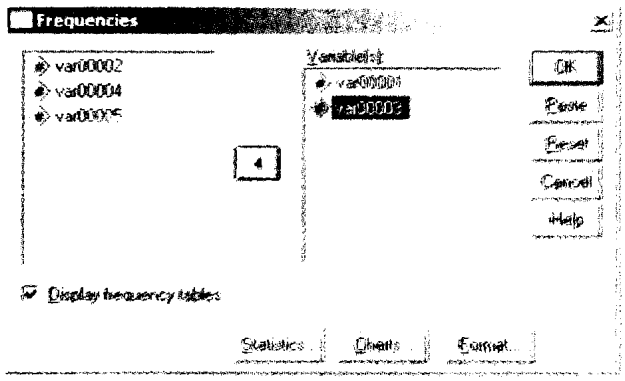

Figure 14.4 : Variables in the active file

Run the Procedure and Examine the Output: After completing the selection process for the procedure and the variables, execute the SPSS command. Most of the commands are executed by clicking OK on the dialog box. The processor will execute the procedures and produce a report in the Output Navigator.

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

14.7 DEFINING, EDITING AND ENTERING DATA

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

Starting the SPSS Session

Click the Start button and select SPSS 11.5 from program menu or double click the icon of SPSS 11.5. When you start an SPSS session, the Data Editor opens automatically. The Data Editor provides a spreadsheet for entering and editing data and creating data files. (see the Figure 14.5).

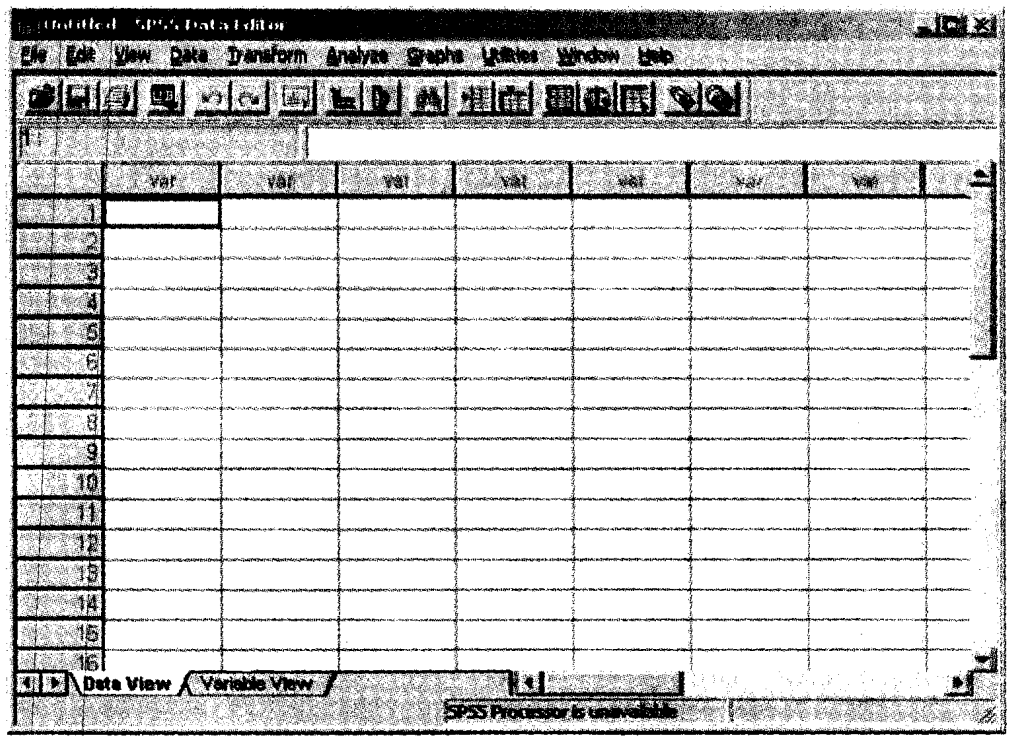

Figure 14.5: Data editor

Important features of the Data Editor include :

- The data is arround in the form of rows and columns in the data editor under
- Rows represent cases. For example, each respondent (student/subject) is a case
- First column represents case numbers
- Columns represent variables. For example, each question is a variable (sometimes it may represent more that one variable)
- Cells represent values. Each cell in defined as the intersection of a row and a column and refers to the value of a particular variable for a specfied case/ direction.

Coding of Data

Before we enter data, we assign codes to the values of variables to make data entry easier. For example, in the "attitude study" gender is a variable that can take on two values. These have been coded so that "1" represents "Male" and "2" represents "Female".

A Sample Code Book is illustrated herewith :

Variable Name : V1
Variable Label : Gender
Variable Labels : Male = 1
Female = 2
Variable Name : V2
Variable Label : Level of Education
Value Labels : Literate = 1
Primary = 2
Secondary = 3
Graduate = 4

Define Variable

Once You prepare your Code Book, you need to include it in the programme for further action to be taken. The process of including the code book is known as "Define Variable".

- A name for the variable (up to 8 characters only)
- A description (label)
- A series of labels which explain the values entered (value labels)
- A declaration as to which values are non-valid and should be excluded from the statistical analysis and other operation (missing values). This information is important to understand the no-response pattern and also to specify the observation which should be excluded from the analysis.

Table 14.2 : Variable definition table provides an example of the above description

Variable Name	Variable Lable	Value Lables	Missing Values	Variable Type
STID	Student identification number	None	None	Number, 6 digits no decimal place
Name	None	None	None	String, 24 character long
Gender	Sex of respondent	M male F female X Unknown	X	String 1 character long
MTL	Marital status	1 Married 2 Widowed 3 Divorced 4 Separated 5 Never married 9 Missing	9	Number, 1 character
DOB	Date of birth	None	None	Date, dd mm yy

To define variable click at Variable View (see the Figure 14.6).

In the left bottom corner of the Data Editor there are two Commands namely Data View and Variable View.

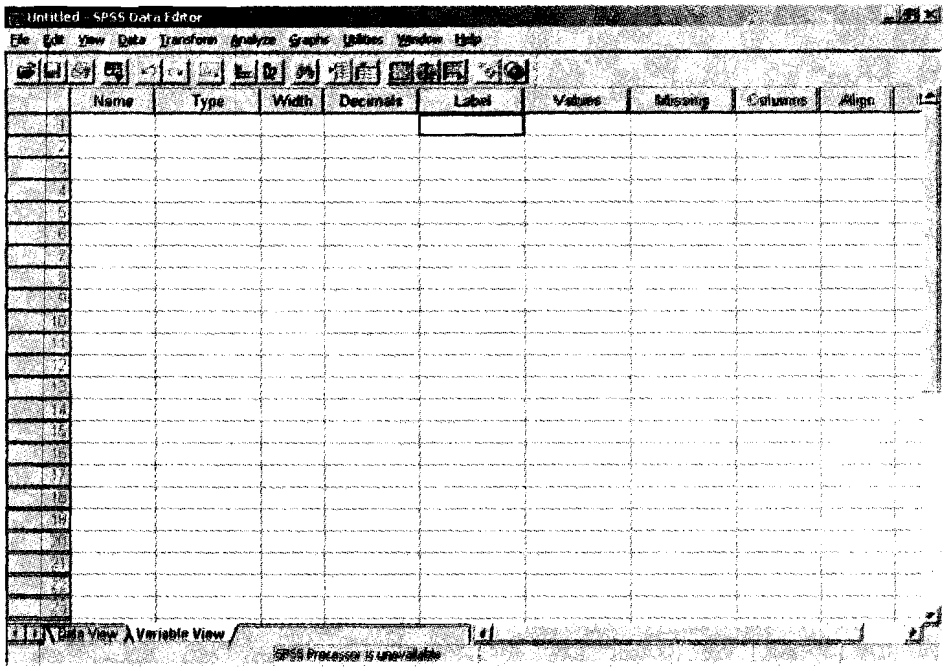

Figure 14.6: Variable view

Next, Click in the first column and type variable name, label & values.

Enter the name you wish to use for the cell. In the example we have chosen the name 'V1' to stand for Gender. Next, Click on the label cell than type "Gender" in the variable label cell. Then click on the values cell, a dialog box will appear (refer to Figure 14.7) in that type value & Value label and then click on Add button.

For example, type "1" in the value box, then click the value label box and type "Male" and finally click the "Add" button. Repeat for the other value of the variable. Once you have finished assigning value labels, click on continue button.

- In case you need to change the labels you can always return to this dialog box. The change button can be used to change a value label.
- The remove button can be used to remove a value label.
- The cancel button can be used to cancel your labeling work and help button can be used to access the SPSS on line help.

Now go ahead and define the remaining variables.

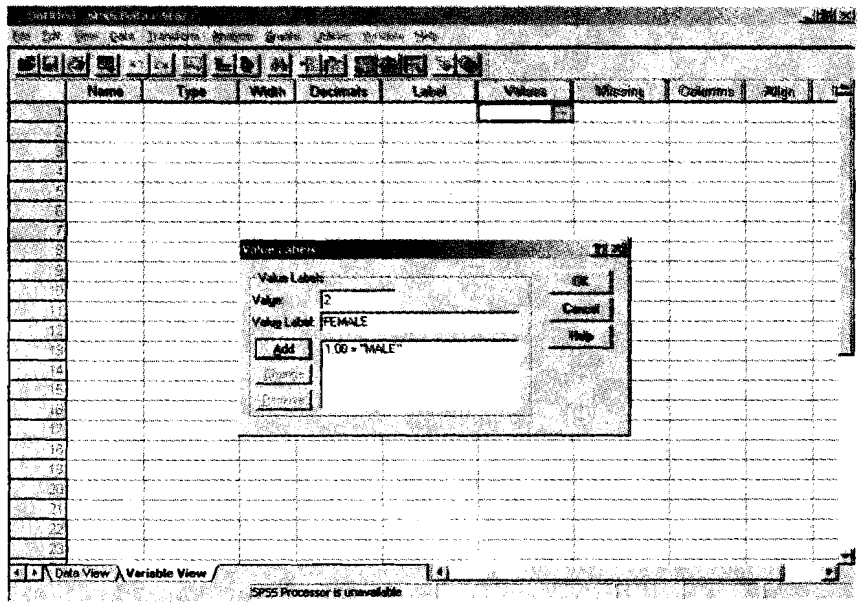

Figure 14.7: Value labels dialog box

So you have seen that defining variables is easy in SPSS. The variable names can be changed and altered with ease even during analysis. Any change made to the working files will be permanently changed only when the data file is saved using 'save' or 'save as' command.

Data can be entered directly using SPSS Data Editor window. However, if the data is large, you are advised to use a data entry package. The data can also be edited/ changed in the data editor window. To change the value in any cell, bring the cursor to the particular cell, enter the new value and press enter. New variables can also be added and the existing variables can be deleted in the Data Editor Window. Let us learn how to enter data using SPSS data editor.

Entering Data

1. Select a cell in the Data Editor
2. Enter the data value. The value is displayed at the top of the Data Editor
3. Press Enter or select another cell to store the value.

Example

To enter the data for the "Attitude Study", simply move the cursor to the upper-left-hand corner and enter 1 for the first respondent's gender (male), then move the cursor one cell to the right and enter 1 for the level of education (literate), and so on. On the screen you will see like Figure 14.8.

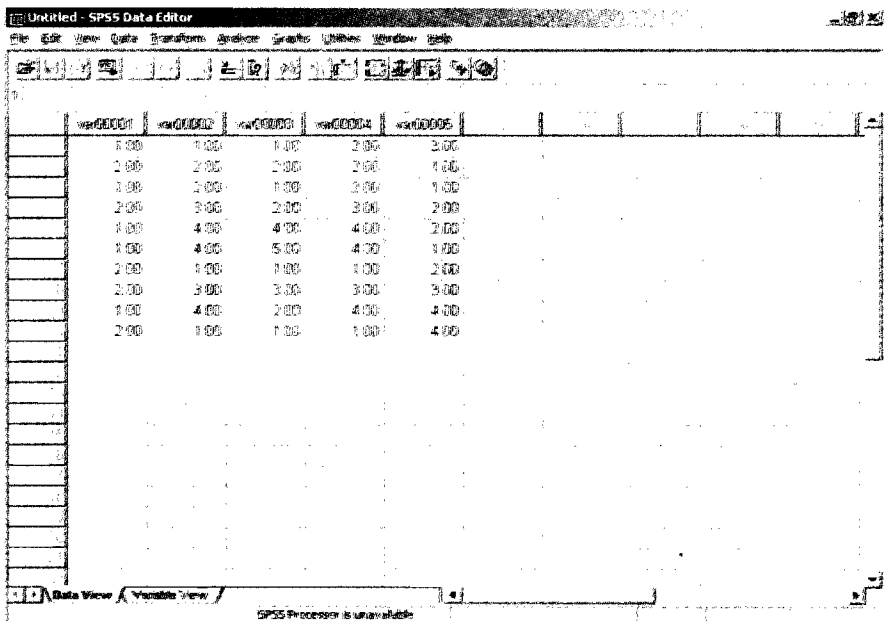

Figure 14.8: Data file

Go ahead and enter the first 10 cases. Now save the data.

Saving the data,

To save data

- From the menus choose:

File

Save

(**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 **OK** button. 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 DATA FILE MANAGEMENT FUNCTIONS

SPSS is very flexible as far as management of data files is concerned. While only one file 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, 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).

14.8.1 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 variables: Refer to Figure 14.9. Adding variables is useful when two data files contain the information about same case but on different variables. For example, the teachers database may contain two files, one having the educational qualifications and the other having the names of the courses taught. Both the files could be combined to analyze the variables available in them. The data on a key and unique variable from both the files can be combined easily. The key variables must have the same name in both the data files. Both the data files should be sorted on the common key variable.

Adding cases: Refer to Figure 14.9. This option is used when the data from two files having 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 data can be merged to create a centralized database by using Add cases command.

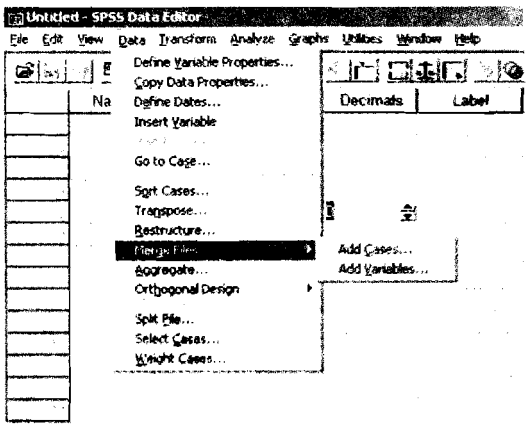

Figure 14.9 : Merge files

14.8.2 Aggregate Data

Aggregate Data command combines groups of cases into a single summary case and creates a new aggregated data file. Refer to Figure 14.10. Cases are aggregated, based on the value of one or more grouping variables. 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 sex, state and region.

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

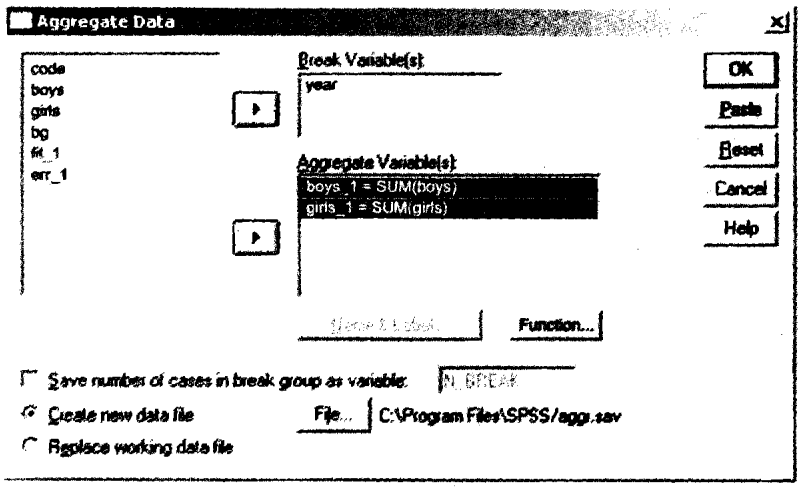

Figure 14.10: Aggregate data file

14.8.3 Split File

The researcher is often interested in the comparison of a summary and other statistics based on certain group behaviour. For example, in a study of learning achievement, the researcher may be interested in comparing the mean scores *for* students belonging to different sex groups. The sex is taken as a grouping variable. Multiple grouping variables can also be selected. 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. The split file is available under Data menu *for* making such comparisons. Refer to Figure 14.10 above.

14.8.4 Select Cases

Select case command can be used *for* selecting a random sub-sample or sub-group of cases based on a specified criteria that includes variables and complex expressions. The following criteria are used for Select Case command.

Select if (condition is satisfied) variable value and their range. Date and time range Arithmatic expression, Logical expression, Functions, Row numbers.

Following the Select Case command, the unselected cases can either be deleted or temporarily filtered. Deleted cases are removed from the active file and cannot be recovered. You should be careful while selecting Delete option. Filtered option will be deleted temporarily. When the Select Case option is on, it is indicated in the Data Editor window.

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

14.9 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.

14.9.1 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 approximate value 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 used for 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 Box-plot. 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.

14.9.2 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. Some of the functions are detailed below:

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.

Recode variables

Recoding of variables is an important characteristics of data management using SPSS. Many continuous and discrete variables need to be recoded for meaningful analysis. Recoding can be done either within the same variable or a new variable can be generated. Recoding in the same variable will replace the original values for this purpose. Recoding in a new variable will replace the old values with new values. The following example illustrates the need and use of recoding variables.

A survey of the primary schools was conducted in Delhi. Alongwith other variables, information on the type of management was also collected. The management code was designed as follows:

1) Government
2) Local bodies
3) Private aided
4) Private unaided
5) Others

Let us assume that a comparative analysis of the government and the private management schools is to be undertaken. This will be done by combining categories 1 and 2 and also 3 and 4. This can be achieved by recoding the management code as 1 (for 1 and 2 categories) and 2 for 3 and 4 categories into a new variable.

Assuming that a database on primary schools in Delhi is available, the enrolment analysis could be attempted by making suitable categories, i.e. schools with less than 50 students, 51 - 150, 151 - 250 and more than 250 students. This could be achieved by recoding the enrolment variable into a new variable 'category'. The analysis could be attempted by changing the class range for category. If at a later stage in the analysis, it is found that a new category is to be introduced, it can again be achieved by recoding the enrolment data.

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 value of count indicating the number of books bought.

Procedure to run count command

Chose Transform from the main Menu

Chose count

Enter the name of a target variable (variable where the count value will be stored)

Select two or more variables of the same type (numeric or string)

Click define variable and specify which value(s) to be counted.

Click OK after the selection has been made.

In survey on learners' achievement, the answer code to each question in language and mathematics could be recorded for each student. The codes could be '1' for the correct answer '2' for the wrong answer and '3' for no reply. Count command can then be used to count the number of correct answers.

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 right hand panel of the Figure 14.11.

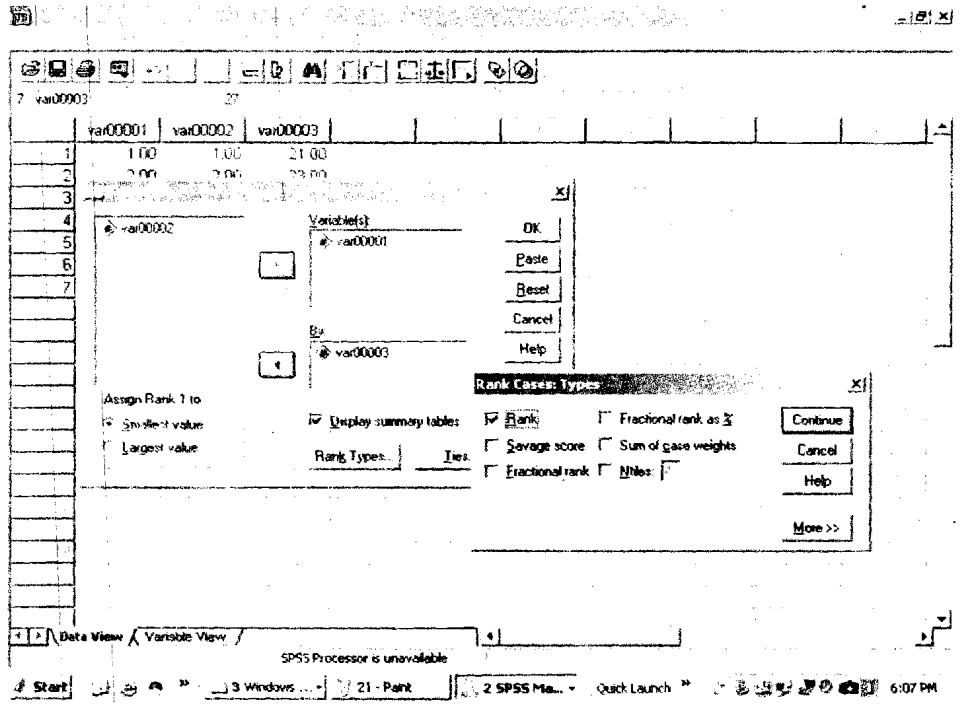

Figure 14.11: Rank case file

Next, we shall review the graphical presentation of data.

14.9.3 Graphical Presentation of Data

SPSS offers extensive facilities for viewing the data and its key features in high resolution charts and plots. From the main menu, select Graphs and the screen shown in Figure 14.12 appears. Various types of Graph that can be drawn using SPSS are indicated in the sub-commands.

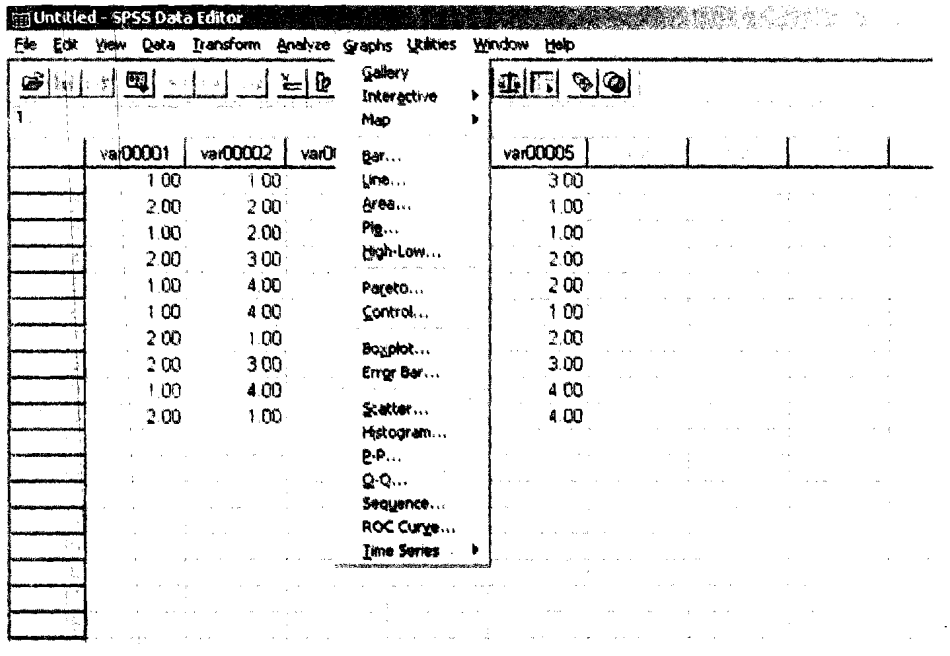

Figure 14.12: Graphics command

Select a chart type from the Graphics menu. This opens a chart dialog box as shown in Figure 14.13.

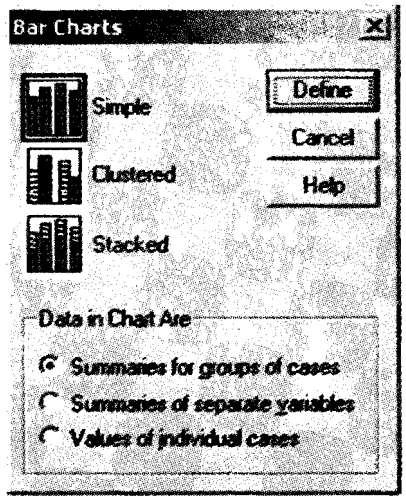

Figure 14.13: Bar chart windows

After the appropriate selections have been made, the output is displayed in the output Navigator window. The chart can be modified by a double click on any part of the chart. Some typical modifications include the following :

- Edit axis titles and labels and footnotes
- Change scale (X - Y)
- Edit the legend
- Add or modify a title
- Add annotation
- Add an outer frame

Another important category of charts is High-Low which are often used to represent variables like maximum and minimum temperature in a day. Stock market behaviour or other similar variables.

Box-plot and Error Bar charts help you to visualize distribution and dispersion. Box plot displays the median and quartiles and special symbols are used to identify outliers, if any. Error Bar chart displays the mean and confidence intervals or standard errors: To obtain a box-plot, choose Box plot from the Graphs menu. The simple box plot for mean scores obtained in English and Hindi is shown in Figure 14.14.

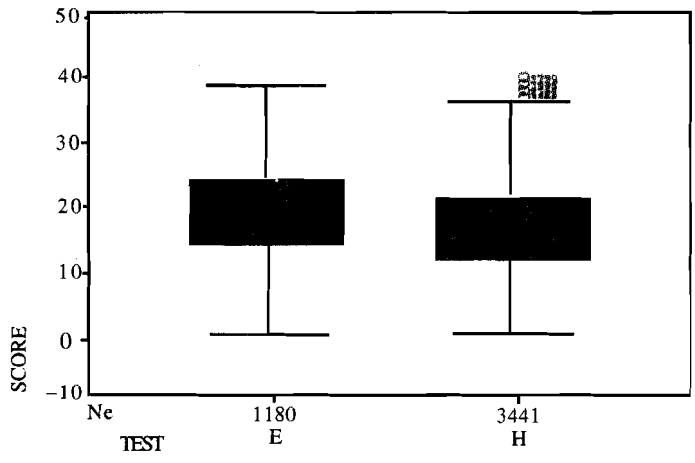

Figure 14.14: Box plot

The above figure shows that there were a large number of outliers in the case of Hindi scores as compared to English. The outliers were along the higher side. This shows that many students were scoring very high marks. The size (numbers) of cases are shown along the X-axis. The boxes show the median and the quartile values for both the tests.

Scatter plots highlight the relationship between two quantitative variables by plotting the actual values along X - Y axis., The scatter plots are useful to examine the actual

nature of relationship between these variables. This could be either linear or non-linear in form. To help visualize the relationship, you can add a simple linear or a quadratic regression line. A 3-D scatter plot adds a third variable in the relationship. You can rotate the two dimensional projection of the three dimensions to delineate the underlying patterns. In order to obtain a scatter plot, select Scatter from the Graphs option.

A *histogram* will be obtained by selecting *Histogram option* from the *Graphs menu*. The variable for which a histogram is to be obtained should be selected from the dialog box. The normal curve can also be displayed alongwith the histogram to visually see the extent of similarity between the actual distribution of values and the normal curve.

Pareto and Control charts are used to analyze and improve the quality of an ongoing process. You may refer to the SPSS manuals for use of these techniques.

14.10 UNDERSTANDING RELATIONSHIPS 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?"

Now, how to analyze this. Next, we shall review some of the various data analysis features specific to parametric and non-parametric tests.

14.10.1 Parametric Test

Under this sub-section we shall review frequency tables, cross-tabulations, correlations, ANOVA, simple regression. We begin with grequency distribution.

A. Freqeuncy Tables

To Analyze the data
From menu, choose

Analyze → Descriptive Statistics → Frequencies

Refer to Figure 14.15.

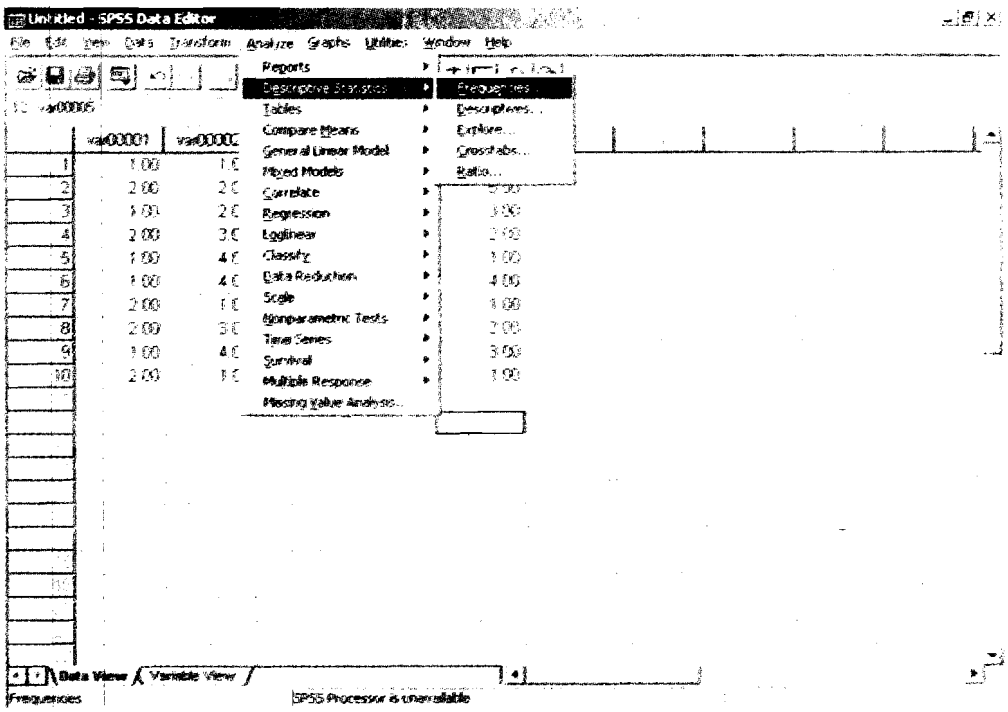

Figure 14.15: Frequencies window

Click the button with the picture of the right arrow. This will move list of selected variables on the right (see Figure 14.16).

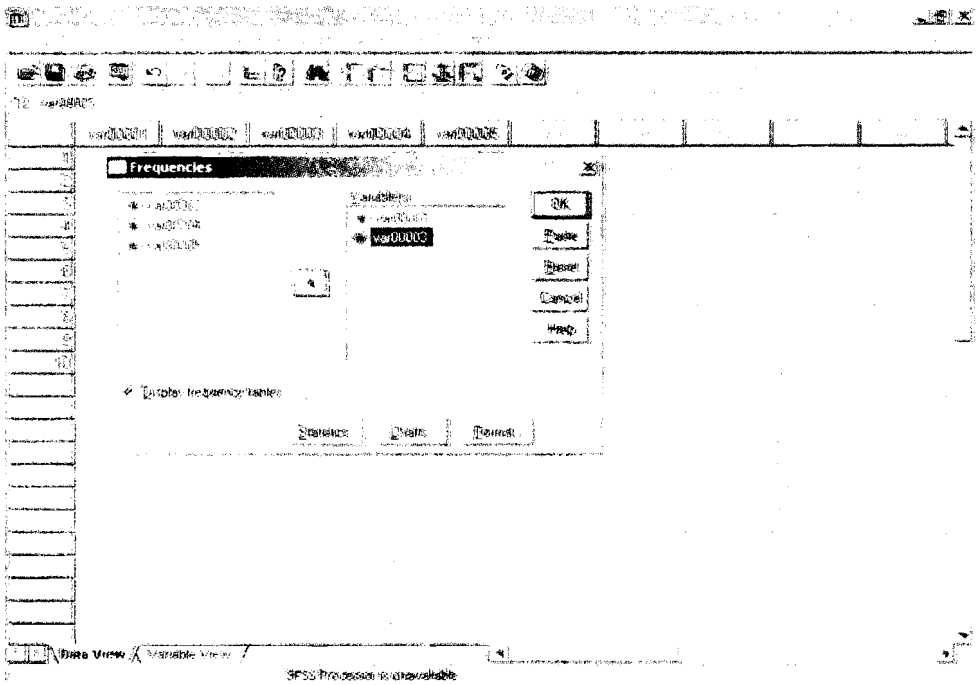

Figure 14.16: Frequencies dialog box

Select on or more variables : To do this click the variable "V1" to select it for analysis, then optionally, you can:

- **Click Statistics** for descriptive statistics for quantitative variables.
- **Click chart** for bar chart, pie-charts and histogram.

If you click statistics you will get a dialog box as shown in Figure 14.17.

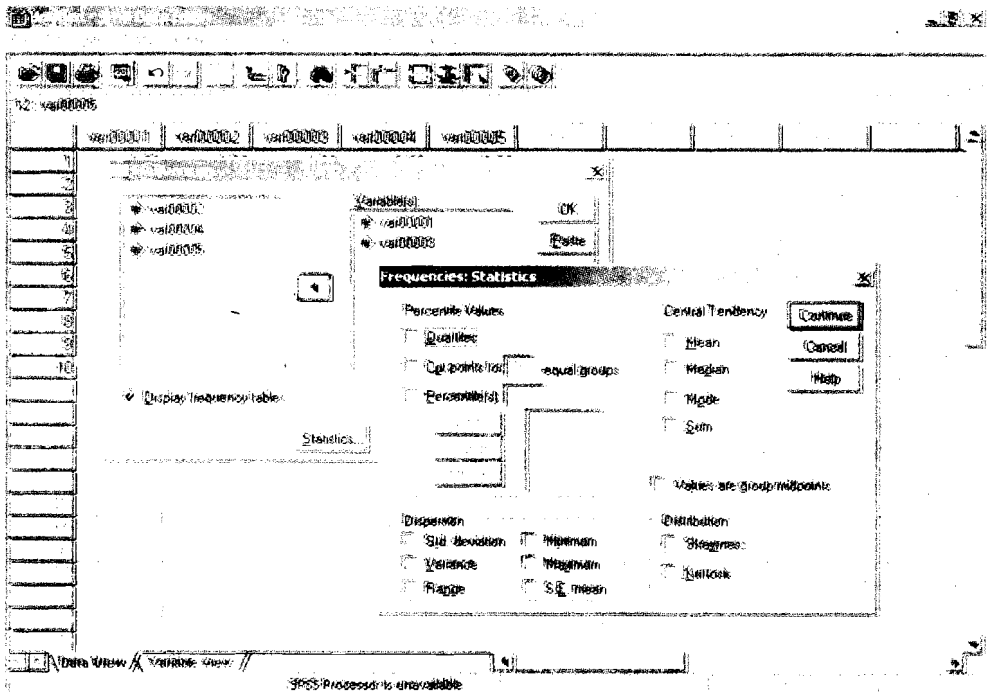

Figure 14.17 : Frequencies statistics dialog box

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 a Output – SPSS Viewer Window showing the Frequencies Tables as illustrated in Figure 14.18.

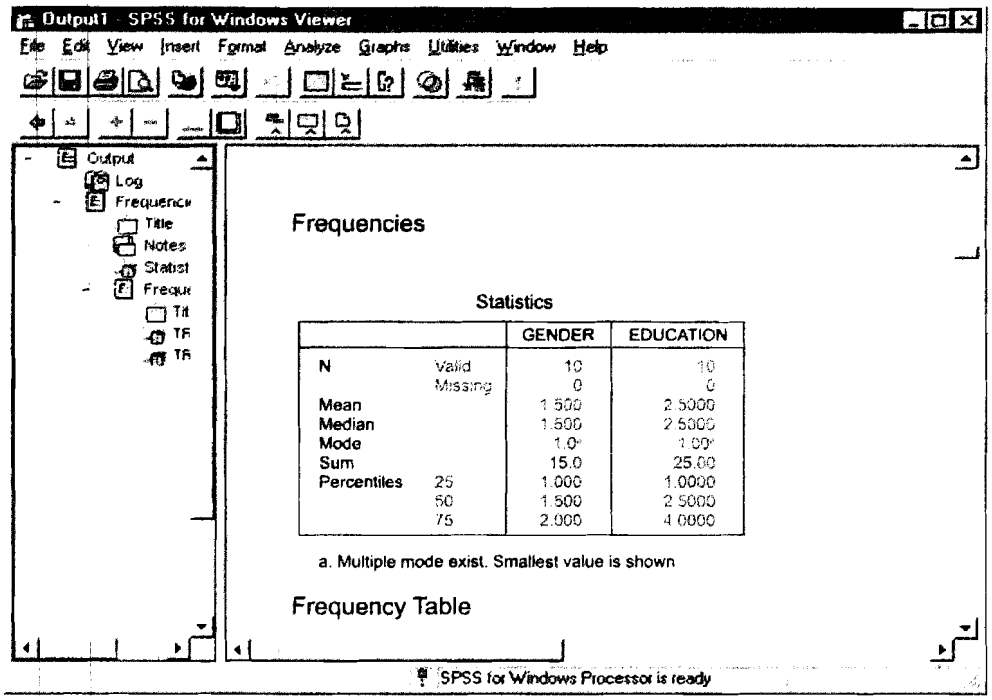

Figure 14.18: Showing frequency table

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.

Nest, let us focus on cross-tabulation feature.

B. Cross-tabulations

Cross-tabulation is the simplest procedure to describe a relationship between two or more categories of variables.

Suppose, you are interested in knowing whether there is an association between two categorical variables such as "gender" and "attitude towards infant feeding" you have to cross-tabulate the two variable and use some statistical tests.

To cross-tabulate:

From the menus (refer to Figure 14.5) choose:

Analyze

Descriptive Statistics

Crosstabe...

You will get a dialog box as shown in Figure 14.19.

Select "Gender" for the row variable and "Attitude" for the column variable. Select one or more row variables and one or more column variables. Optionally you can:

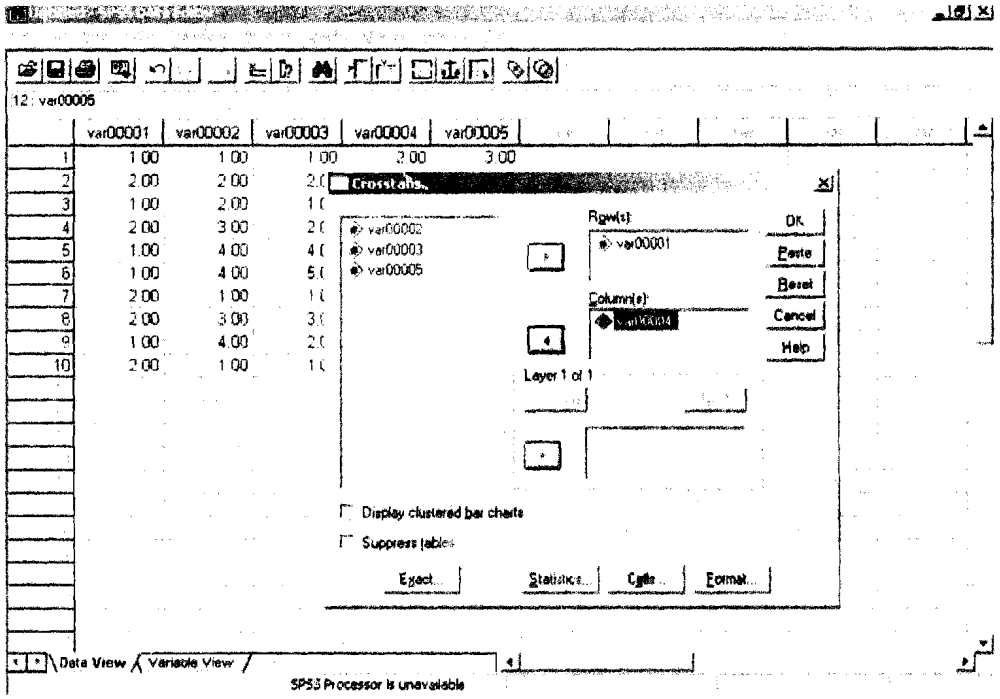

Figure 14.19: Crosstabs dialog box

- Click Statistics for statistical tests
- Click cells for percentages

If you click 'statistics' you will get statistics dialog box as shown in Figure 14.20.

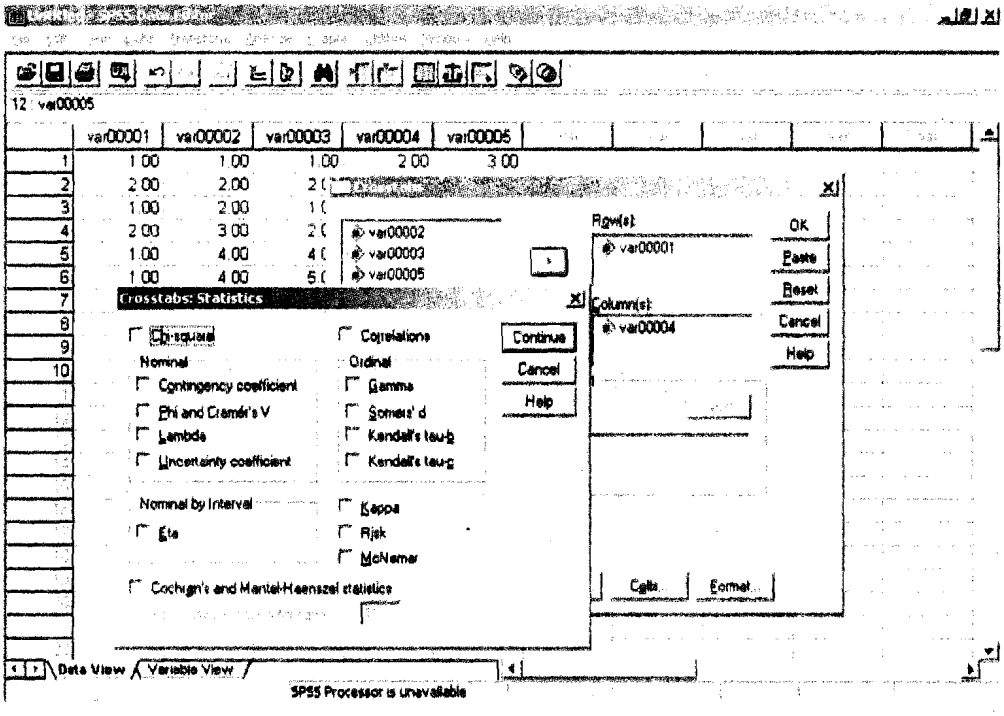

Figure 14.20: Statistics dialog box

In this dialog box click the box next to the statistical tests you wish to apply. Say for example if you wish to apply chi-square test, click the box next to chi-square. Then click the continue button. Next, click the cells button. Click on the row or colum (or both) percentage box. Click the continue button, the click the OK button. The table shown in Figure 14.21 will appear in the output navigator window:

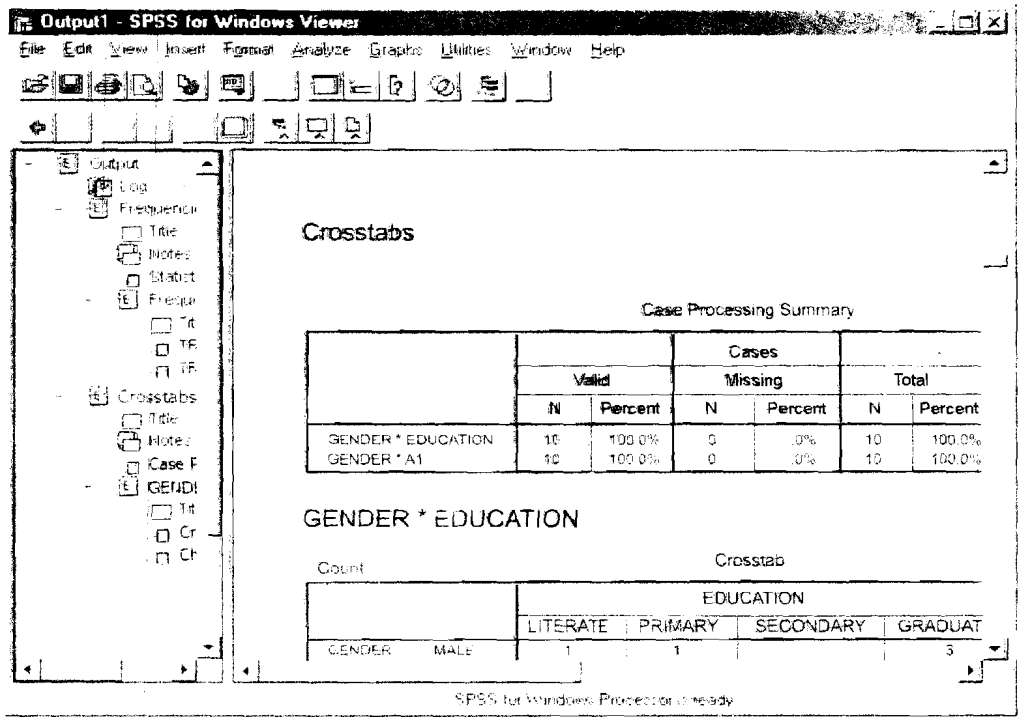

Figure 14.21: Showing crosstab table

Next, let us review bivariate correlation.

C. Bivariate Correlations

To obtain Bivariate Correlations

From the menus (refer to Figure14.15), choose:

Analyze

Correlate

Bivariate

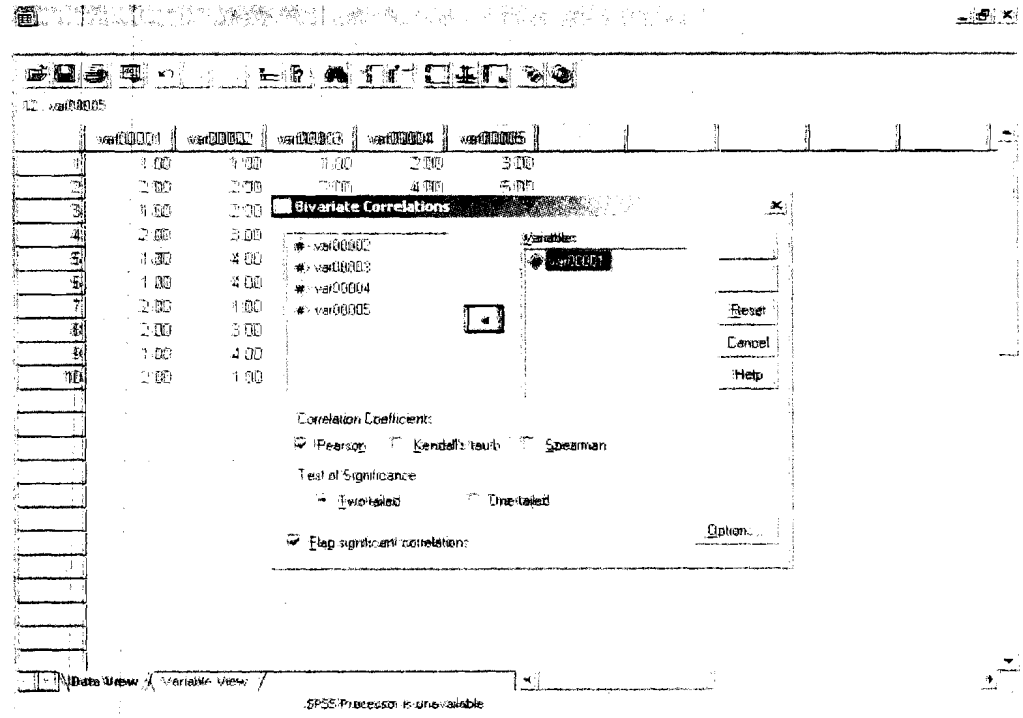

Figure 14.22: Bivariate correlations

Select two or more numeric variables. The following options are also available:

- Correlation coefficients
- Test of significance
- Flag significant correlation

Next, we shall review how to calculate the independent sample T-test.

D. Independent Samples T-test

To obtain an Independent samples T-test

From the menus (refer to Figure 14.23) choose:

Analyze

Compare Means

Independent-Samples T-test

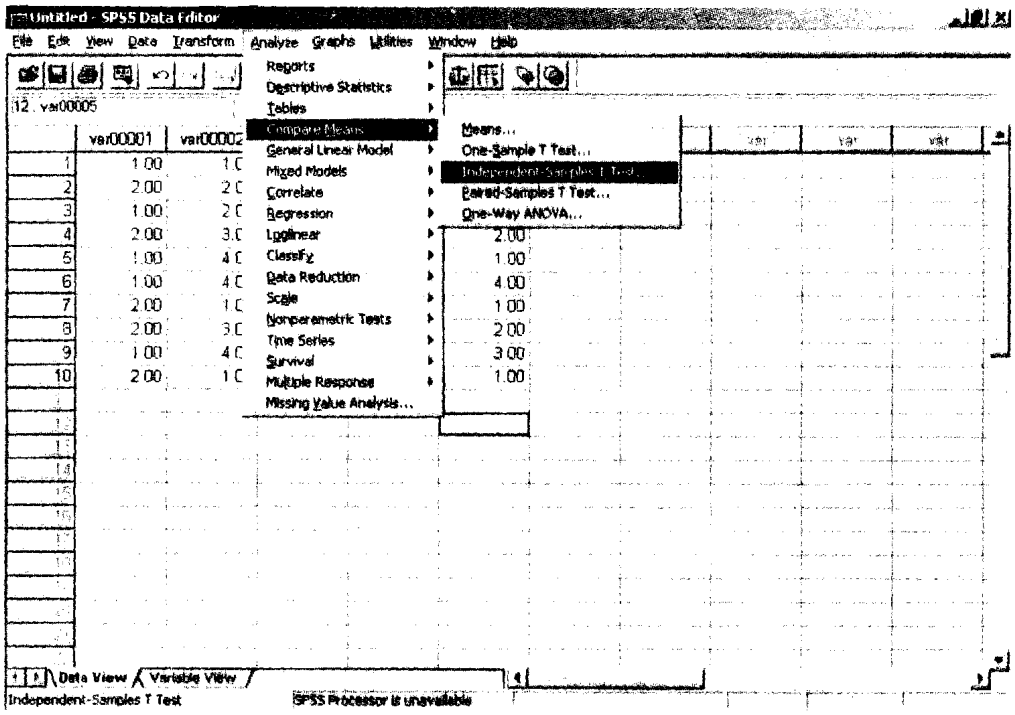

Figure 14.23: Independent samples T-test dialog box

First select a quantitative test variable then select a single grouping variable and click Define Groups to specify two codes for the groups you want to compare.

For example; you wish to test a hypothesis that "Did males and females have similar mean attitude scores? To test the hypothesis of equality of means for two groups we can use the t-test statistic.

Figure 14.24 displays the independent samples T-test

- Select "Attitude" as the test variable
- Select "Gender" as the grouping variable
- Click the button define groups
- Enter "1" for group 1 and "2" for group 2 (as shown in Figure 14.24)
- Click the continue button to return to the previous dialog box.
- Click 'OK' button to run the procedure.

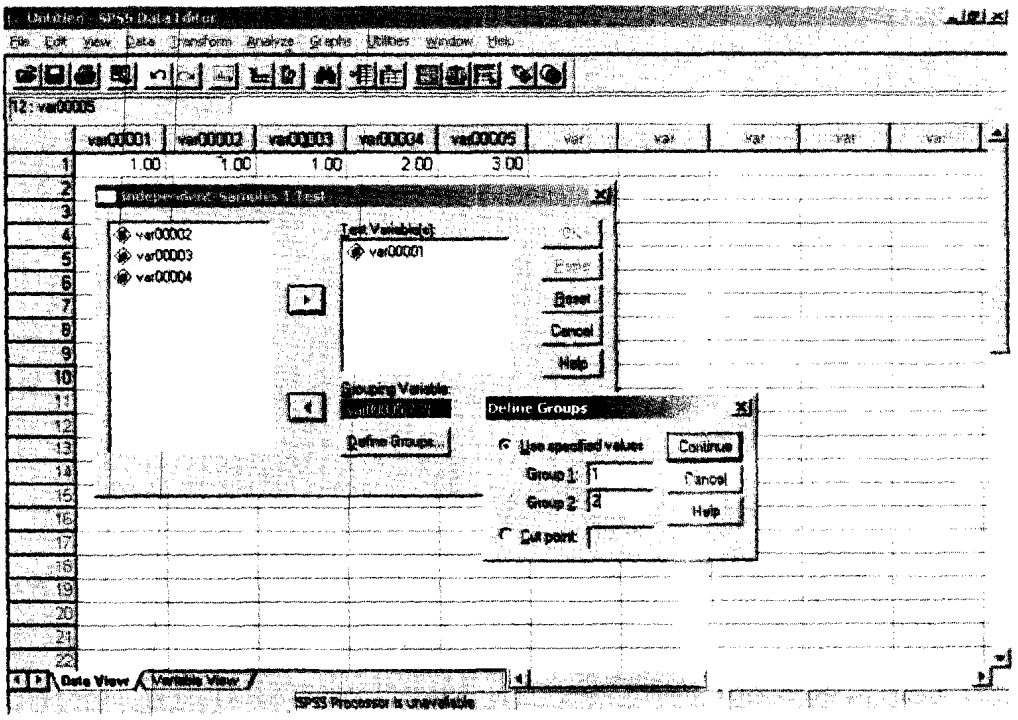

Figure 14.24: Define groups dialog box

The results shown in Figure 14.25 will appear in the output navigator window.

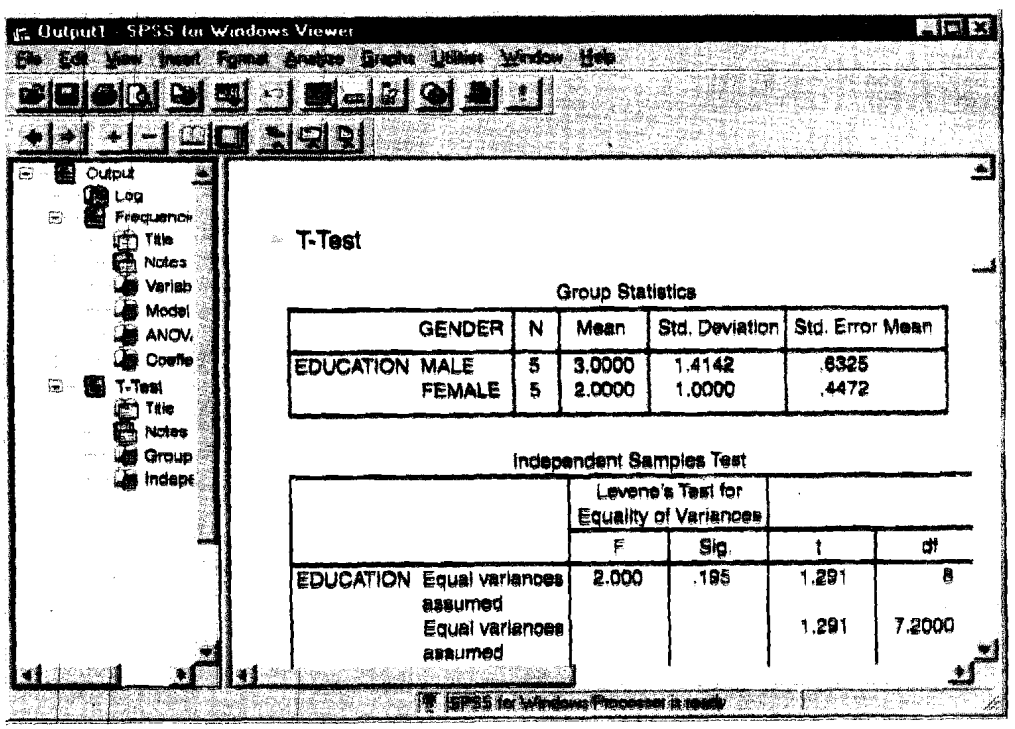

Figure 14.25: Showing T-test tables

Comparing several means (ANOVA)

When we are interested in an independent variable that has more than two groups, then we will need to use the analysis of variance (ANOVA).

Suppose, you are interested in testing the hypothesis, "Do students in each of the three groups of religious affiliation have similar mean attitude scores?"

From the menu (refer to Figure 14.26) choose

Analyze → Compare Means → Oneway ANOVA

You will see the dialog box, as shown in Figure 14.26.

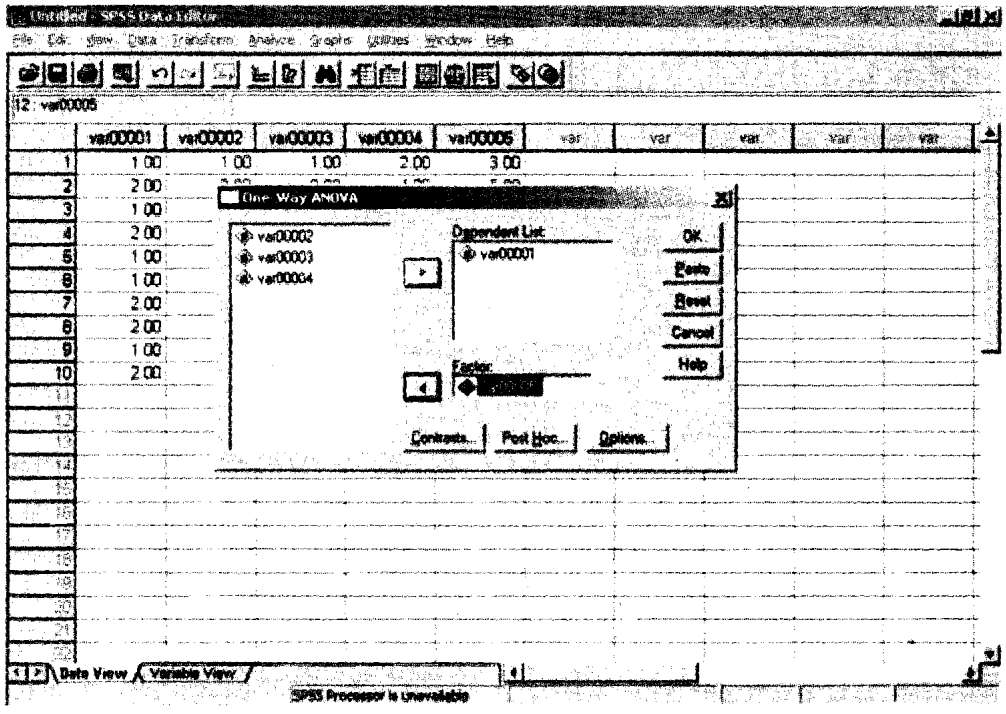

Figure 14.26: Oneway ANOVA dialog box

- Select attitude ( ) as the dependent variable and religious affiliation as the factor (i.e., independent variable)
- Click the 'OK' button to run the procedure. You will see the results illustrated in Figure 14.27 in the output navigator window.

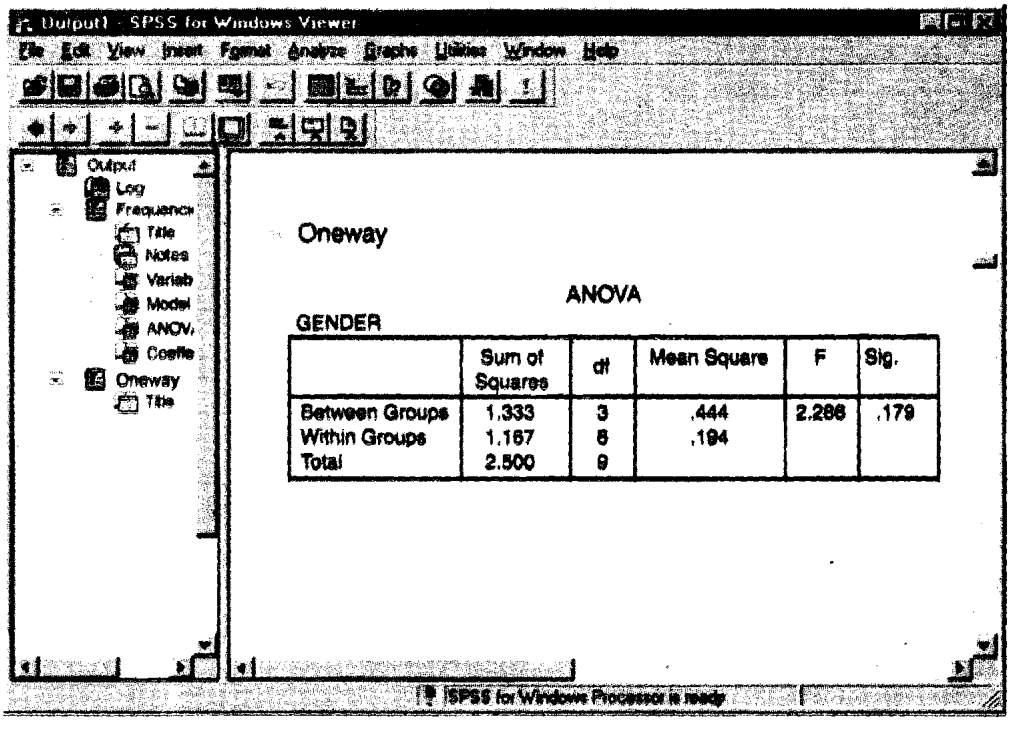

Figure 14.27: Showing oneway ANOVA tables

Next, we shall review simple regression process.

E. Linear Regression

Suppose, you have a question. "How well can we predict 'attitude' of students if we know something about their levels of education?" We need to conduct a simple regression analysis to answer the question.

To conduct this analysis, from the Analyze pull-down menu, select Regression, then choose Linear...The dialog box shown in Figure 14.28 will appear.

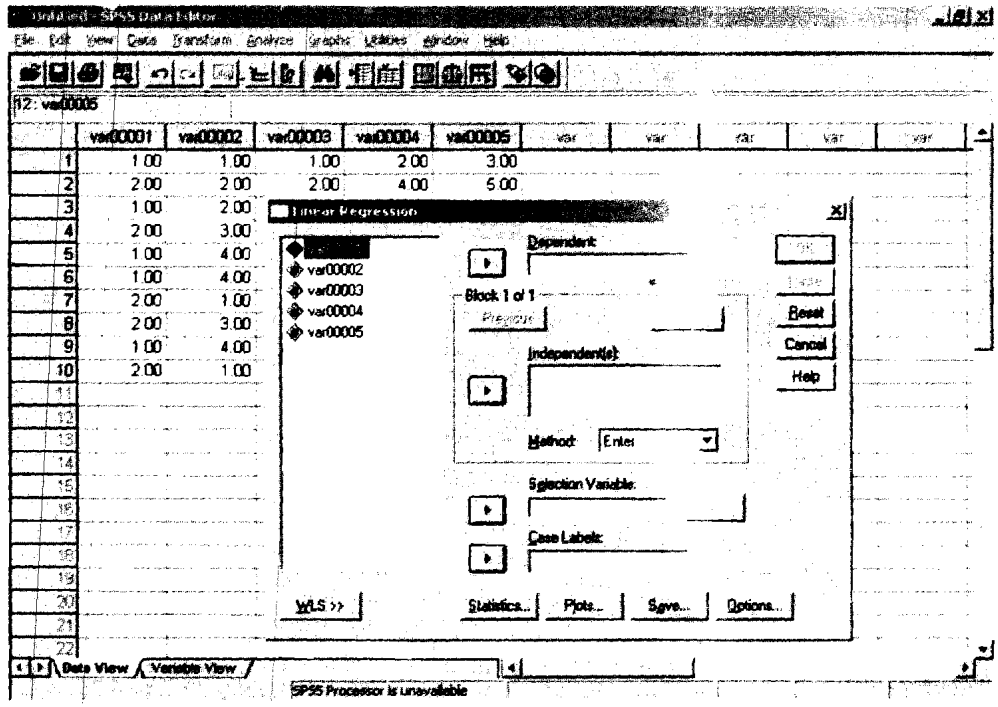

Figure 14.28: Linear regression dialog box

Select 'attitude' as the dependent variable and students levels of education as the independent variable. Note that SPSS provides for many important options that are useful in conducting regression analysis. These are available via the Analyze..., Plots..., Save..., and Options... buttons. Readers interested in learning more about regression analysis are encouraged to review Schroeder, Sjoquist, and Stephan (1986), as well as, the chapter on regression in the SPSS manuals (which details these analysis options). Click the OK button to run the procedure. The results of regression analysis will appear in the output navigator window as shown in Figure 14.29.

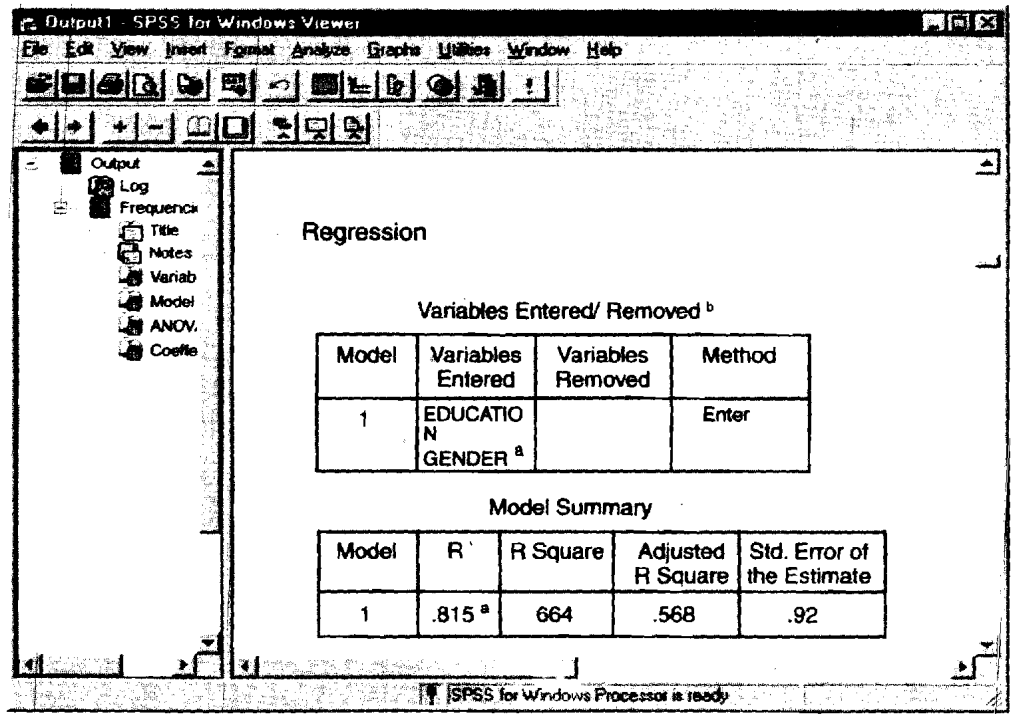

Figure 14.29: Showing regression results

Linear regression is the most commonly used procedure for the analysis of a cause and effect relationship between one dependent variable and a number of independent variables. The dependent and independent variables should, be quantitative. Categorical variables like sex and religion should be recoded to dummy (binary) variables or other types of contrast variables. An important assumption of the regression analysis is that the distribution of the dependent variable is normal. Moreover, the relationship between the dependent and all the independent variables should be linear and all observations should be independent of each other.

SPSS provides extensive scope for regression analysis using various types of selection processes.

The method of selecting of independent variables for linear regression analysis is an important choice which the researcher should consider before running the analysis. You can construct a variety of regression models from the same set of variables by using different methods.

You can enter all the variables in a single step or enter the independent variables selectively.

Variable selection method is shown in Figure 14.30.

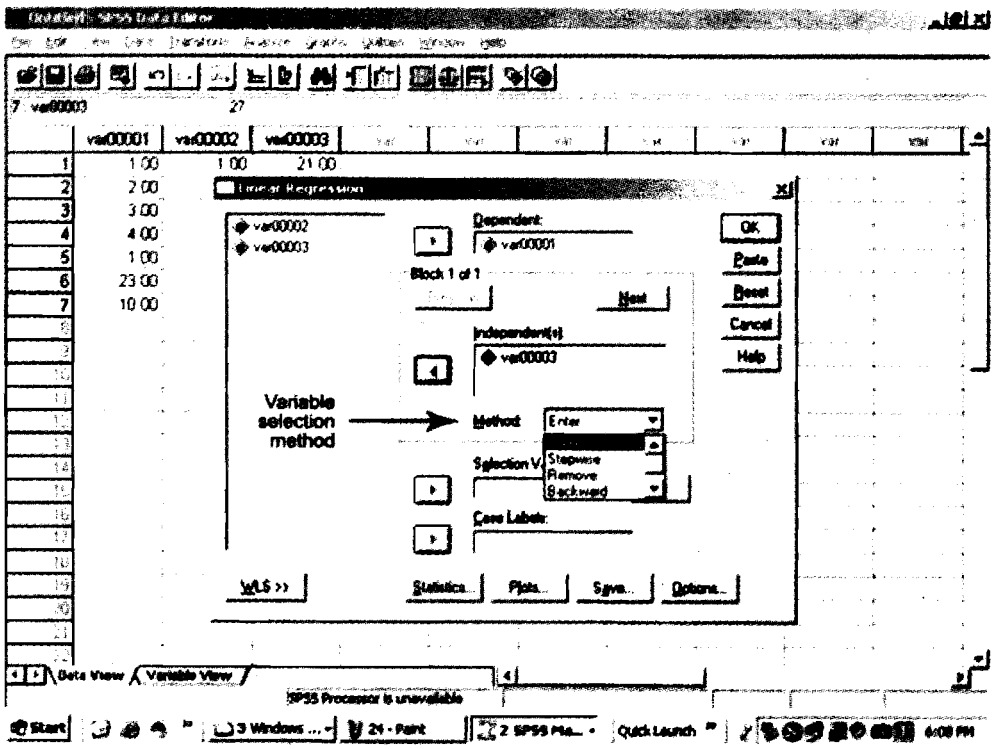

Figure 14.30: Variable selection method

It allows you to specify how independent variables are entered into the regression analysis. The following options are available:

- Enter: To enter all the variables in a single step, select Enter option.
- Remove: To remove the variables in a block in a single step.
- Forward: It enters one variable at a time based on the selected criterion.
- Backward: All variables are entered in the first instance and then one variable is removed at a time on the selected criterion.
- Stepwise: Stepwise variable entry and removal examines the variables in the *block* at each step for entry and removal. This is a forward step procedure.

All the variables must pass the tolerance criterion to be entered in the equation; regardless of the entry method specified. The default tolerance limit is 0.0001. A new variable will not be entered if it causes the tolerance of another variable already entered to be dropped below the tolerance limit.

Linear Regression Statistics

The following statistics (refer to Figure 14.31) are available on linear regression models. Estimates and Model Fit are the two options which are selected by default.

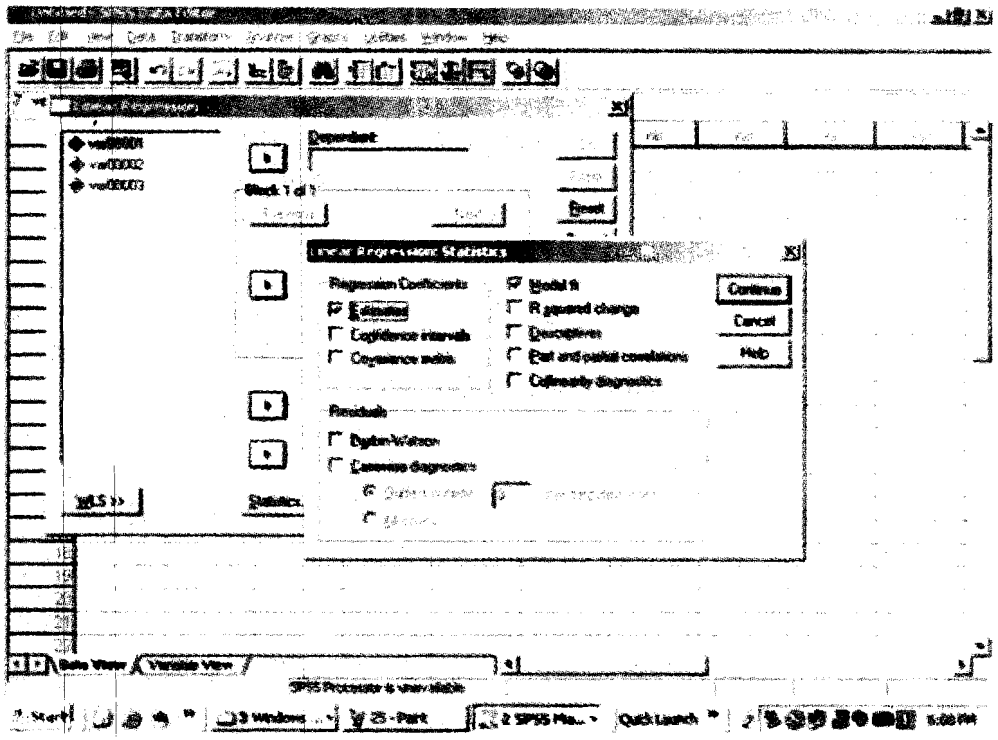

Figure 14.31: Linear regression statistics

Regression coefficients: The Estimates option displays regression coefficient, *B*, standard error, standard coefficient beta, t-value; and two tailed significance level of *t*. Covariance matrix displays a variance covariance matrix of regression coefficients with covariance off the diagonal and variance off the diagonal. A correlation matrix will also be Displayed.

Model fit: The variables entered and removed from the model are displayed. Goodness of fit statistics, R-square, multiple R, and adjusted R-square, standard error of the estimate and an analysis of variance table is displayed.

If other options are ticked, the statistics corresponding to each of the options are also displayed in the Output Navigator. If the data does not show linear relationship and the transformation procedure does not help, try using Curve Estimation procedure.

14.10.2 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
- Run Test
- One sample Kolmogorov Semonov 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.

Chi-Square

Chi-square test (refer to Figure 14.32) 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.

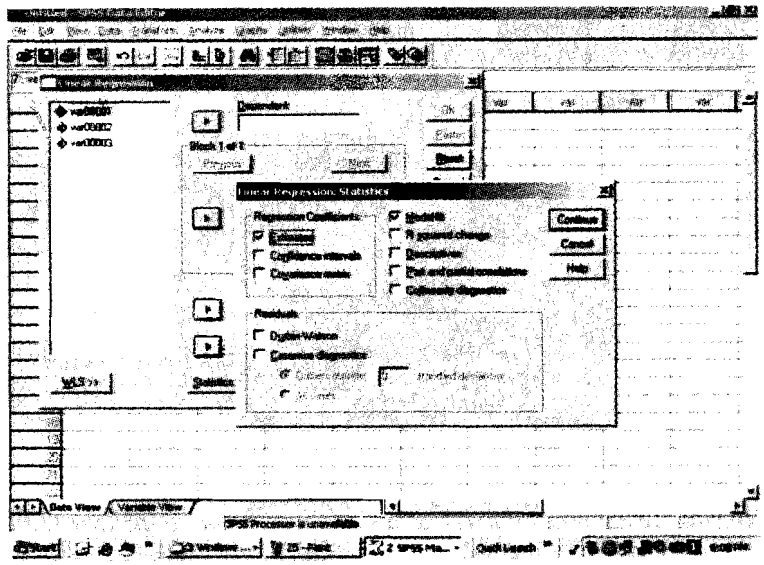

Figure 14.32: Chi-square test

Consider that a bag contains red, white and yellow balls. You want to test the hypothesis that the bag contains all type of balls in equal proportion. To obtain Chi-square test, choose Chi-square from Non-parametric tests in the Statistics command. Select one or more variables. Each variable produces a separate output.

By default, all categories have equal expected values as shown in the figure above. Categories can have user specified proportions also. In order to provide user specific expected values, select the Values option and add the user expected values. The sequence in which the values are entered is very important in this case. It corresponds to the ascending order of the category values of the test variable.

14.11 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 last command. Production mode is useful if you run the same set of time-consuming analysis periodically.

The SPSS Production facility uses command syntax file to tell SPSS about the commands to be executed. We have already discussed the important features of the command syntax. 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. Start SPSS Production program from the start window of window 95/98 or later version. Specify the syntax file that you want to use in the production job. Click Browse to select the Syntax File. Save the production file job. Run the production file job at any time.

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

14.12 STATISTICAL ANALYSIS SYSTEM (SAS)

Like the SPSS, the Statistical Analysis System (SAS) package calculate descriptive statistics of your choice e.g., Mean, Standard Deviation etc. SAS is available for both main frame and personal computers. It is strong in its treatment of data, in clarity of its graphics and in certain business applications. The various statistical procedures carried out by SAS are always preceded by the word PROC which stands for procedure. The most commonly used SAS statistical procedures are as follows: (Sprinthall et.al, 1991).

- PROC MEANS: Descriptive statistics (mean, standard deviation, maximum and minimum values and so on).
- PROC CORR: Pearson correlation between two or more variables.
- PROC t-TEST: t-test for significant difference between the means of two groups.
- PROC ANOVA: Analysis of variance for all types of designs (one way, two-way and others).
- PROC FREQ: Frequency distribution for one or more variables.

As pointed out by Klieger (1984) SAS package is comparatively more difficult to use due to its procedural complexities. For greater details on SAS package you are advised to consult the books by Klieger and Sprinthall.

14.13 NUDIST

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 invaluable aid in research.

NUDIST (Non-numerical unstructured data indexing, searching and theorizing) 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:

1) 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.

2) 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 counselors. The researcher can create a node in NUDIST as 'Role of Academic Counselors'. Researcher will select text in the transcripts where learners have talked about this role and merge it into role of Academic Counselors. Information can be retained in this node and researcher can take print in different ways in which learners talk about the role of academic counselors.

3) Crossing themes: Taking the same example of role of counselors, the researcher can relate this node to other nodes. Suppose the other node is qualifications of counselors. There are two categories like Graduate and Post Graduate. The researcher will ask NUDIST to cross the two categories, role of counselors and

qualification of counselors to see for example whether there is any relation between graduate counselors and their role than the post graduate counselors and their role. NUDIST software generates information for a matrix with information in the cells reflecting different perspectives.

4) 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.

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

For further details on NUDIST software you may like to consult the following:

Kelle. E.(ed.), *Computer aided qualitative data analysis,* Thousand Oaks, CA: Sage, 1995.

Tesch, R., *Qualitative research: Analysis types and software tools,* Bristol, PA: Falmer, 1990.

14.14 LET US SUM UP

The foregoing details examined the various types of statistical application of the SPSS in data management, presentation and analysis. The discussion was based on the assumption that you have a basic understanding of the statistical methods. It was highlighted that the researchers must try to explore the data using various simple but powerful statistical techniques. In this connection, six characteristics of the data were examined for exploring it fully. The procedures involved in the use of various statistics were also discussed in detail. Procedures for running regression analysis to understand the relationships between variables were also discussed. Those of you who are comfortable with the basic statistical procedures in SPSS can explore the advanced features, including those aimed at automating the statistical analysis using the SPSS Production facility and also the use of scripting in data analysis.

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	1551	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	2.00	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**	**2.00**	**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 V: 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
32804	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	49144	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	[illegible]304	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	77524	20675
13263	92240	89559	66545	06433	38634	36645	22350	81169	97417
66309	31466	97705	46996	69059	33771	95004	89037	38054	80853
348	05291	38713	82303	26293	61319	45285	75784	50043	44438

om pp.99 and 100 of The Rand Corporation, *A Million Random Digits with 100,000 Normal iates*, The Free Press, Glencoe, Ill., 1955, with permission of the Rand Corporation.

(Continued)

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	06668	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	95322	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