UNIT 9 STATISTICAL PACKAGES

Structure

9.0 OBJECTIVES

After reading this Unit, you would be able to:

- define the word "statistics";
- describe the different data types and data formats;
- illustrate the role of statistical methods;
- describe the features of statistical packages like SPSS and MS-Excel; and
- portray some of the popular statistical packages.

9.1 INTRODUCTION

One of the features of the development of modern world is the development of the capacity to convert observations in numbers. The science, which deals with numbers, is statistics. It crunches the numbers and organises them in a meaningful way so that information is generated. This information builds up knowledge and thus the development goes on. Advances in computing have come handy in this as they help in doing this part of job accurately, timely effectively and convincingly.

Computer can help immensely in the statistical analysis. There exist numerous statistical tools and the need is to identify their actual usage. Even with the use of a statistical package many statistical procedures require a lot of prior knowledge and insight.

In this Unit, we have tried to build-up a case for the usage of statistics by first defining statistics and what it can do to your data. You will further find the definition of the data types. An explanation of the common tasks that are performed in a preliminary analysis is also given. This Unit also presents a description of two popular packages (MS Excel and SPSS) and gives a glimpse of some other statistical packages.

9.2 STATISTICAL PACKAGES

Before going on to the definition of statistical packages, one needs to revisit the definition of statistics and its functions. In this section, we would highlight the areas/ problems that statistics as a discipline addresses to and the kind of data one gets for the statistical applications.

9.2.1 Definition

The term "Statistics" is used as a "collection of numerical facts or data". It is also used in terms of a "body of methods and techniques for analyzing numerical data". Statistical techniques have many purposes, which include methods and procedures for summarising, simplifying, reducing and presenting raw data. It then makes predictions, tests hypotheses and infers characteristics of a population from the characteristics of a sample. In other words, Statistics is generally thought of as serving two functions. One is to describe sets of data; the other is to help in drawing inferences. When you are studying only a sample, there is possibility that your assumption may not be accurate and you can never be certain that you have drawn the correct inference. For this reason the inferential use of statistics may be thought of as helping you to make decisions under conditions of uncertainty. It is different from guessing, because Statistics also provides you with a method of estimating how reliable your conclusions are. With each statistical statement that you make, you indicate the probability that findings like yours could have been the result of chance factors.

A statistical package is the software for the collection, organisation, interpretation, and presentation of numerical information. The need for a statistical package has arisen because of the complexity of calculations involved in making inferences from the data. The advances in computing technologies have made statistics a yet more powerful field.

According to Ripley (2004), "The most widely used piece of statistical packages/ software for statistics is Excel. SPSS and SAS dominate certain communities, and Minitab is widely used in teaching. Many niche products, e.g. GenStat, Generalised Linear Interactive Modelling Package (GLIM), Stata and S-PLUS dominate the high-end, hence is widely seen in methodology papers."

9.2.2 Data Measurement

Statistical data is generally obtained in many formats such as spreadsheets (e.g. MS Excel) or databases (e.g. MS Access). Data may also be received in various open formats such as typically tab-delimited text (*.dat, *.tab, *.txt), comma-separated text or fixed-width text data (*.dat, *.txt).

The data could be of two types, qualitative and quantitative. Most of the statistical methods are based on the quantitative data. Quantitative variable is a variable whose values are numbers with real numeric meaning. It consists of mainly two types of data viz. discrete and continuous. A set of data is said to be discrete if the values/observations belonging to it are distinct and separate, i.e. they can be counted e.g. number of books in a library. Whereas, a set of data is said to be continuous if the values/observations belonging to it may take on any value within a finite or infinite interval.

There are four well-known levels of measurement scales i.e. nominal, ordinal, interval, and ratio. There is a relationship between the level of measurement and the appropriateness of various statistical procedures. For example, it would be impractical to compute the mean of nominal measurements. Data must be measured on an interval or a ratio scale for the computation of means and other statistics to be valid. Therefore,

if data are measured on an ordinal scale, the median but not the mean can serve as a measure of central tendency. Let us have a brief discussion of what these scales are:

1) **Nominal Scale**

Nominal measurement consists of assigning items to groups or categories. No quantitative information is conveyed and no ordering of the items is implied. Nominal scales are therefore qualitative rather than quantitative. Religious preference, race, and sex are all examples of nominal scales. Frequency distributions are usually used to analyse data measured on a nominal scale. The main statistic computed is the mode. Variables measured on a nominal scale are often referred to as categorical or qualitative variables. Nominal variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories

2) **Ordinal Scale**

Measurements with ordinal scales are ordered in the sense that higher numbers represent higher values. However, the intervals between the numbers are not necessarily equal. Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do not allow us to say "how much more."

3) **Interval Scale (Cardinal Scale)**

On interval measurement scales, one unit on the scale represents the same magnitude on the trait or characteristic being measured across the whole range of the scale. Interval scales do not have a "true" zero point, however, and therefore it is not possible to make statements about how many times higher one score is than another. True interval measurement is somewhere between rare and nonexistent in the behavioral sciences. A good example of an interval scale is the Fahrenheit scale for temperature. Interval variables allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them.

4) **Ratio Scale**

Ratio scales are like interval scales except they have true zero points. Ratio variables are very similar to interval variables. In addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus they allow for statements such as x is two times more than y. A typical example of ratio scales is measure of time or space. Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

9.2.3 Functions of Statistics

Many people intend to use statistical techniques in their research. It is definitely a good practice to substantiate your claims with the help of data. Statistics has various functions, which can be broadly categorised as follows:

1) *Summarise and Describe data:* One summarises and describes the data in order to view data at a glance. If it is nominal or ordinal data, one makes cross-tabulations and graphs; if it is interval or ratio data then z-scores are calculated.

2) *Variance and distribution of the data:* In order to measure the spread of the data and knowing its distributions one makes tables and charts and graphs for nominal/ordinal data and histograms with normal curve or box plots with inter-quartile range for interval/ratio data.

3) *Compare groups:* When one has to compare two or more populations then one makes cross-tabulations for nominal/ordinal data and employ testing of hypothesis for continuous/numeric data divided into groups.

4) *Identify relationships:* In order to identify relationships in the data, one uses cross-tabulations for nominal/ordinal data; calculate correlation coefficient and scatter plot for Interval/ratio data or go for linear regression/ ANOVA for data with one dependent and 2 or more predictor variables.

5) *Identify groups of similar cases:* Carrying out hierarchical cluster analysis solves the problem of identifying groups of similar cases or k-means cluster analysis. One uses Discriminant analysis for identify characteristics of known groups.

6) *Identify groups of similar variables:* Factor analysis is carried out to identify groups of similar variables.

Self Check Exercise

1) Define a Statistical package. Describe its need and purpose.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

9.3 FEATURES OF SOME STATISTICAL PACKAGES

Advances in computing especially the advent of the personal computer (PC) have made computing a game of the commoners. Today one has the computing power as one can easily load software of his choice or need into his PC. There is a plethora of read-made computer packages available today. Now one can find different statistical packages for applications to different disciplines. We will describe two such packages that are ready available and are popular and user friendly. We will also give a glimpse of some other packages in the subsequent section.

9.3.1 Microsoft Excel

Microsoft Excel is a big worksheet (it can take data rows in thousands across 256 columns). This worksheet can be used for data entry and for performing calculations by click of buttons. It has a "paste function where you can paste any formula from a big list of inbuilt functions. MS Excel can be used to create tables, and graphs and perform statistical calculations. The work done in MS Excel can be easily copied and pasted to many window-based programs for further analysis.

According to Pottel, "Spreadsheets are a useful and popular tool for processing and presenting data. In fact, Microsoft Excel spreadsheets have become somewhat of a standard for data storage, at least for smaller data sets. The fact that the program is often being packaged with new computers, which increases its easy availability, naturally encourages its use for statistical analysis. However, many statisticians find this unfortunate, since Excel is clearly not a statistical package. There is no doubt about that, and Excel has never claimed to be one. But one should face the facts that due to its easy availability many people, including professional statisticians, use Excel, even on a daily basis, for quick and easy statistical calculations. Therefore, it is important to know the flaws in

Excel, which, unfortunately, still exist today! …. "Excel is clearly not an adequate statistics package because many statistical methods are simply not available. This lack of functionality makes it difficult to use it for more than computing summary statistics and simple linear regression and hypothesis testing". However in MS Excel 2003 aspects of the some statistical functions, including rounding results, and precision have been enhanced.

The MS Excel worksheet is a collection of cells. As we have earlier said, there are 65,000 (rows) X 256 (columns) cells in an MS Excel worksheet. Each row or column can be used to enter data belonging to one category. Data entry in MS Excel is as simple as writing on a piece of paper. MS Excel assigns each column a field depending upon the type of data. It supports various data formats; one can choose a data format by formatting the cells.

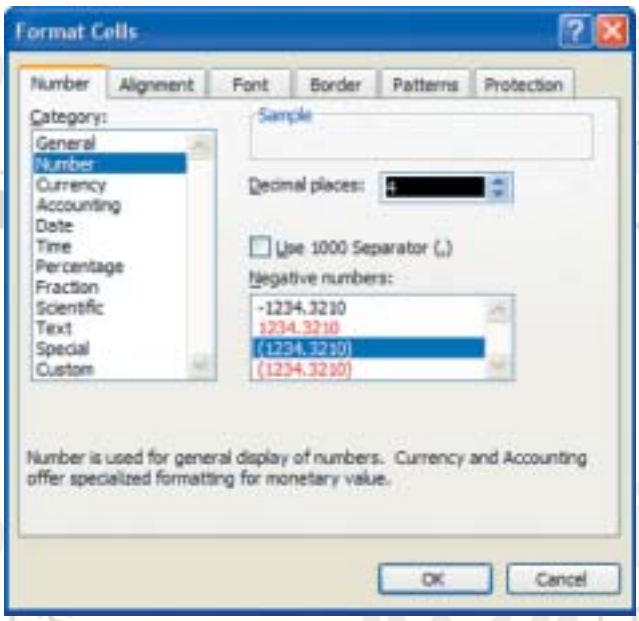

Fig. 9.1: Screen Snapshot of MS-Excel for Choosing Data Format

Once the type of cells is defined it is easy to enter the data without taking care of the format. MS Excel can perform usual calculations on the data so entered. It has an insert function (f_x) icon that contains many inbuilt functions like sum, count, max/min, standard deviation etc. In fact it has a plethora of built-in functions that performs special calculations without even typing the formula. To perform a calculation one has to select a function and specify the range of values on which it has to be applied. These functions are known as paste functions.

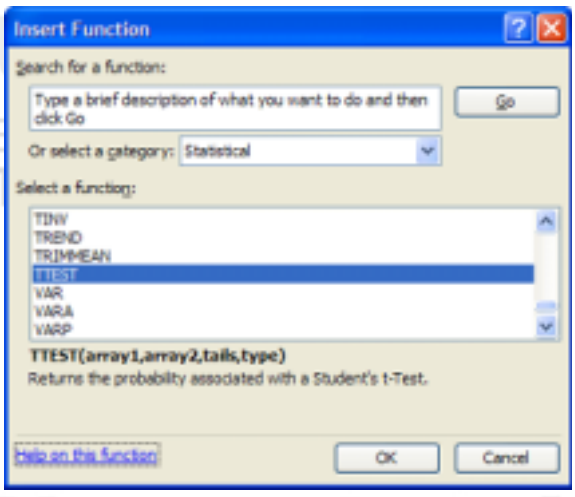

Fig. 9.2: Screen Snapshot of Function Menu Options in MS-Excel

We will concentrate on the statistical functions and see some of the major statistical functions of MS Excel. As you can see in figure 9.2, once you go to the function menu and choose "statistical" category, you will be asked to select a function. Suppose you have chosen t-test. You will be told on the same screen that t-test returns the probability associated with a student's t-test. Now if you are still not comfortable with the description, you may select help on this function, which is at the bottom left of the screen. More help is offered in the following form.

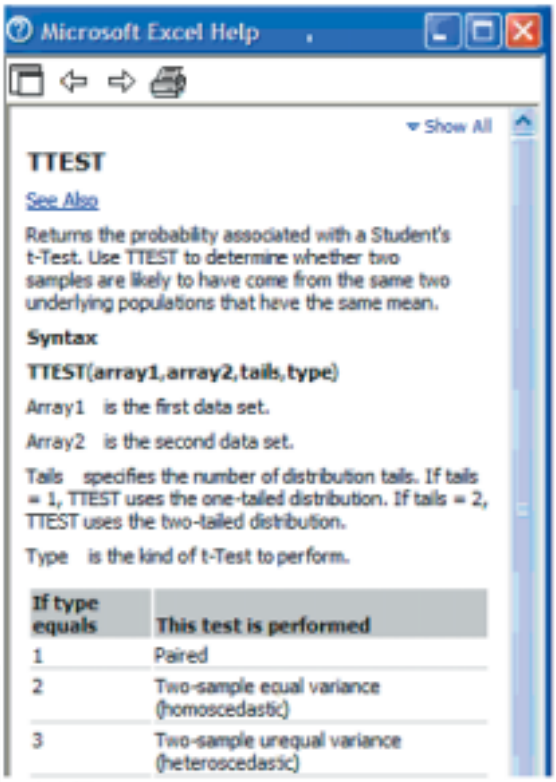

Fig. 9.3: Screen Snapshot of Help for TTest in MS-Excel

MS Excel has a built-in statistical package for taking you in further details of data analysis. It provides a set of data analysis tools called the Analysis ToolPak, which you can use to save steps when you develop complex statistical analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

To access these tools, click Data Analysis on the Tools menu.

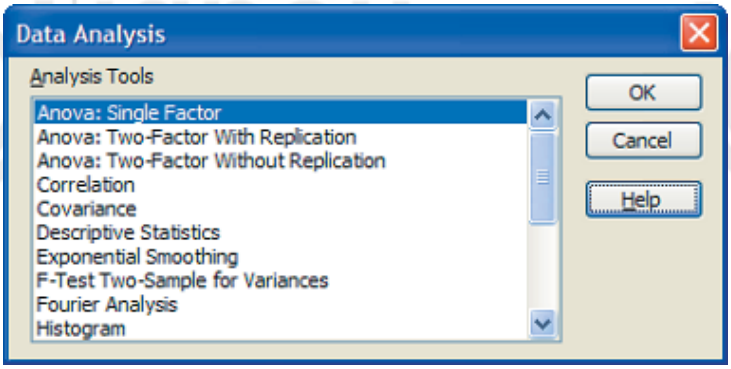

Fig. 9.4: Screen Snapshot of Data Analysis Function in MS-Excel

Let us have a brief description of these tools. The table given below highlights their functions and uses.

Table 9.1: Data Analysis Tools in MS-Excel

Sl. No.	Tool	Function	Use
1	ANOVA	The ANOVA tools provide different types of variance analysis.	Test of the hypothesis that each sample is drawn from the same underlying probability distribution,
2	Correlation Covariance	Calculate the correlation coefficient/ Covariance between two variables when measurements on each variable are observed for each of N subjects.	Examine each pair of measurement variables to determine whether the two measurement variables tend to move together
3	Descriptive Statistics	Generates a report of univariate statistics for data, providing information about the central tendency and variability of your data.	Describes the data in an interpretable format and sow summary statistics like mean, mode, median, std. deviation, skewness, kurtosis and range etc.
4	Exponential Smoothing	Predicts a value based on the forecast for the prior period, adjusted for the error in that prior forecast	Forecast on the basis of a smoothing constant.
5	F-Test	Performs a two-sample F-test to compare two population variances.	Test that these two samples come from distributions with equal variances.
6	Moving Average	Projects values in the forecast period, based on the average value of the variable over a specific number of preceding periods.	Forecast trends on the basis of past figures.
7	Regression analysis	Performs linear regression analysis by using the "least squares" method to fit a line through a set of observations.	Analyse how a single dependent variable is affected by the values of one or more independent variables.
8	Sampling analysis	Creates a sample from a population by treating the input range as a population.	Infer about a population on the basis of a sample.
9	t-Test, z-Test	Determine whether the two samples are likely to have come from distributions with equal population means	Compare two population means when the population variances are known and unknown

Source: Based on the Microsoft Office Excel 2003 help function

You have seen that MS Excel can do virtually most of common statistical calculations. There are two more features that are worth mentioning when one talks about the statistical functions of MS Excel. These two are cross tabulations, pivot tables and the graphical features.

MS Excel can be used to create cross tabulations or two-way frequency tables across categorical variables. In MS Excel there is a pivot table wizard which helps in creating tables in multi-dimensions. Let us explain these concepts with the help of an example. The data given below is the percentage contribution of a country to world research in a particular subject area.

Table 9.2: Percentage Contribution of a Country to World Research in a Particular Subject

Countries	Chem.	Engg.	Clin. Med.	Phy.	Mat. Sc.	Plant. Sc.
Argentina	0.57	0.26	0.31	0.56	0.36	0.92
Australia	1.56	2.18	2.51	1.36	1.69	4.56
France	6.17	4.9	5.7	7.08	5.81	5.16
Germany	9.38	6.56	8.12	9.89	9.07	6.6
Hungary	0.83	0.44	0.25	0.54	0.41	0.61
India	4.02	2.68	0.85	2.51	4.02	3.49
Ireland	0.22	0.27	0.36	0.21	0.27	0.37
Israel	0.79	1.06	1.26	1.37	0.71	1.11
Italy	3.62	4.02	4.4	4.42	2.48	2.33
Japan	11.33	9.75	8.21	11.14	13.77	6.98

Now suppose you want to make a pivot table that would enable you to visualise a country whose contributions differ in the disciplines of physics and chemistry. You can simply drag the subject field in rows and column. This would enable you to see that e.g. the shaded countries Hungary and India have significantly different contributions.

Count of Countries	Chem.										Grand
Phy.	0.22	0.57	0.79	0.83	1.56	3.62	4.02	6.17	9.38	11.33	Total
0.21	1										1
0.54											1
0.56		1									1
1.36					1						1
1.37			1								1
2.51											1
4.42						1					1
7.08								1			1
9.89									1		1
11.14										1	1
Grand Total	1	1	1	1	1	1	1	1	1	1	10

There could be many more such cross-tabulations depending upon the need of the researcher. The most pleasant part of working with MS Excel is the ease with which you can drag these fields and have a customised layout as per your wish. It is said that in MS Excel, the most demanding work is to input the data, the analysis being the easiest. You can use pivot table effectively to present data where two-dimensional tables are important. One of the major advantages of this feature is that once the table is prepared, we can change the summary from one characteristic to another.

Similarly, if you want to make a graphical presentation of the data then you can go to the chart wizard and choose the chart that you want to make. MS Excel has a built-in facility to create graphs and charts. There are several types of charts and graphs supported by MS Excel like bar charts, line charts, pie charts and scatter diagrams etc. The chart wizard menu can be summoned by clicking on the graph icon from the menu bar. The chart wizard looks like the following

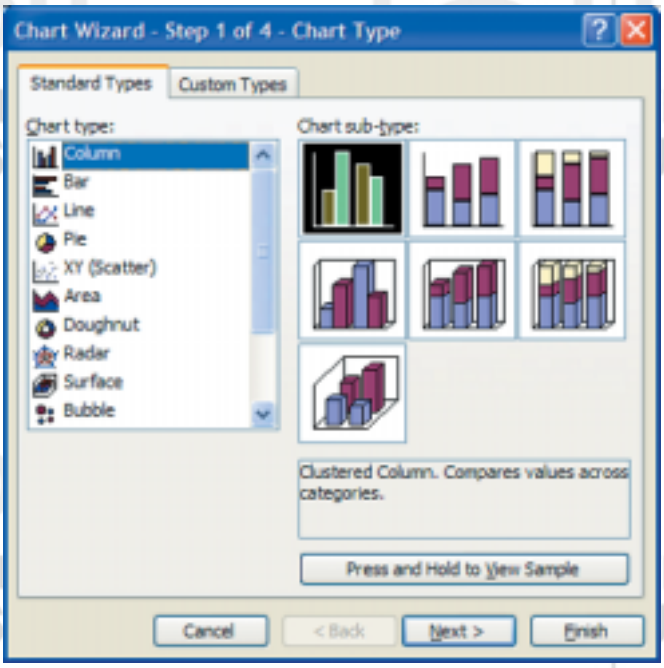

Fig. 9.5 : Screen Snapshot of Chart Wizard in MS-Excel

Now suppose you want to make a pivot table that would enable you to visualize a country whose contributions differ in the disciplines of physics and chemistry. You can simply drag the subject field in rows and column. This would enable you to see that e.g. the shaded countries Hungary and India have significantly different contributions.

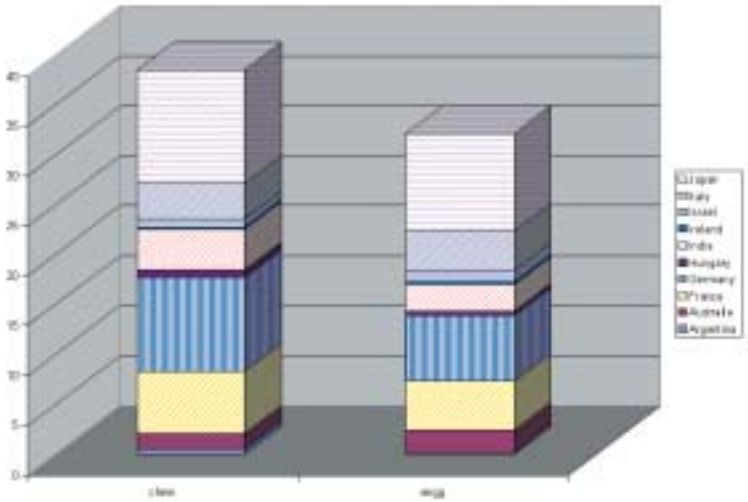

Fig. 9.6 : Screen Snapshot of Chart Companies of the Contribution of Countries to Chemistry & Engineering

So you have seen that it is pretty simple to create a graph and visualize your data once you have some data in your worksheet. These graphs are very useful for common user requirements. We can cut copy and paste these graphs to any document of MS Word or PowerPoint.

Self-Check Exercise

2) Enumerate the data analysis tools in MS-Excel.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

9.3.2 SPSS

SPSS is the short form for Statistical Package for Social Sciences (SPSS). It is a very popular package due to its features and compatibility with other window-based programs. In the late 1960s, three Stanford University graduate students developed the SPSS statistical software system.

SPSS can take data input from many packages like dBase (*.dbf), Excel (*.xls), Lotus 123 (*.w*) and others like *.dat and *.txt. It can filter the data and perform analysis only in selected cases.

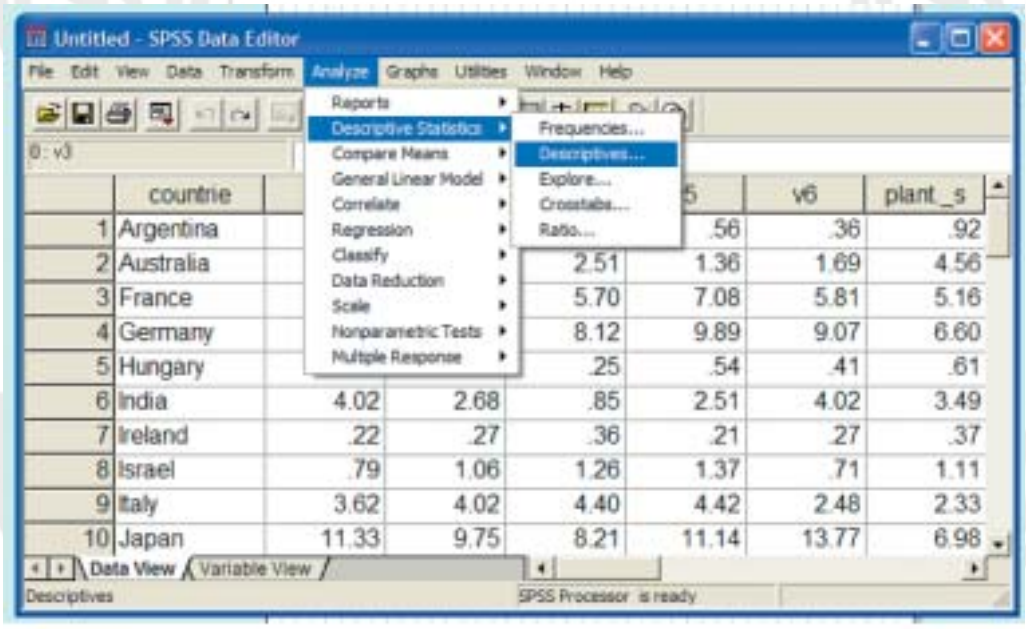

Fig. 9.7 : Screen Snapshot of SPSS Data Editor

Once you open a data file you can go to the analyze menu and start working on the statistical aspects of the data. Figure 9.7 shows you the menu for the descriptive statistics. You can find the frequencies, cross-tabulations, and ratios etc. You can see that there is a long list of statistical analysis in the analyze menu. We will give you a glimpse of what these functions do.

Fig. 9.8: Screen Snapshot of Help Topics in SPSS

Table 9.3 : Functions of Statistical Analysis in SPSS

Tool/ Module	Statistical Procedures
Report	OLAP (Online Analytical Processing) Cubes, Summarize procedure, Report Summaries in rows, Report Summaries in Columns
Descriptive Statistics	Frequencies procedure, Descriptive procedure, Explore procedure, Cross-tabs procedure, Ratio Statistics procedure
Compare means	Means procedure, One-Sample T Test procedure, independent-Samples T Test procedure, Paired-Samples T Test procedure, One-Way ANOVA procedure
General Linear Model	GLM Univariate procedure
Correlate	Bivariate Correlations procedure, Partial Correlations procedure, calculates statistics measuring either similarities or dissimilarities (distances)
Regression	Linear Regression procedure, Curve Estimation procedure
Classify	K-Means Cluster Analysis procedure, Hierarchical Cluster Analysis procedure, Discriminant Analysis procedure
Data Reduction	Factor analysis procedure
Scale	Reliability analysis procedure, Multidimensional scaling procedure
Non Parametric Tests	Chi-Square Test procedure, Binomial Test procedure, Runs Test procedure, One-Sample Kolmogorov-Smirnov Test procedure, Two-Independent-Samples Tests procedure, Tests for Several Independent Samples procedure, Two-Related-Samples Tests procedure, Tests for Several Related Samples procedure
Multiple Response	Define Multiple Response Sets procedure, Multiple Response Frequencies procedure, Multiple Response Cross-tabs procedure.

Source: SPSS for Windows Release 11.0 Help

One can see that many of these tools are available in MS Excel also, but the difference is that the output given by SPSS contains many more details regarding the statistical aspects of the findings. For example the cross-tabs procedure forms two-way and multi-way tables in both MS Excel and SPSS, but in SPSS it also provides a variety of tests and measures of association for two-way tables. The supporting statistics provided in SPSS is Pearson chi-square, likelihood-ratio chi-square, linear-by-linear association test, Fisher's exact test, Yates' corrected chi-square, Pearson's r, Spearman's rho, contingency coefficient, phi, Cramér's V, symmetric and asymmetric lambdas, Goodman and Kruskal's tau, uncertainty coefficient, gamma, Somers' d, Kendall's tau-b, Kendall's tau-c, eta coefficient, Cohen's kappa, relative risk estimate, odds ratio, McNemar test, and Cochran's and Mantel-Haenszel statistics. SPSS is thus more comprehensive.

SPSS also supports several statistical graphs. It displays many statistics on the graph itself. It has a feature that helps you to find a chart that is most suitable for your data, which is called "Chart Galleries by Data Structure".

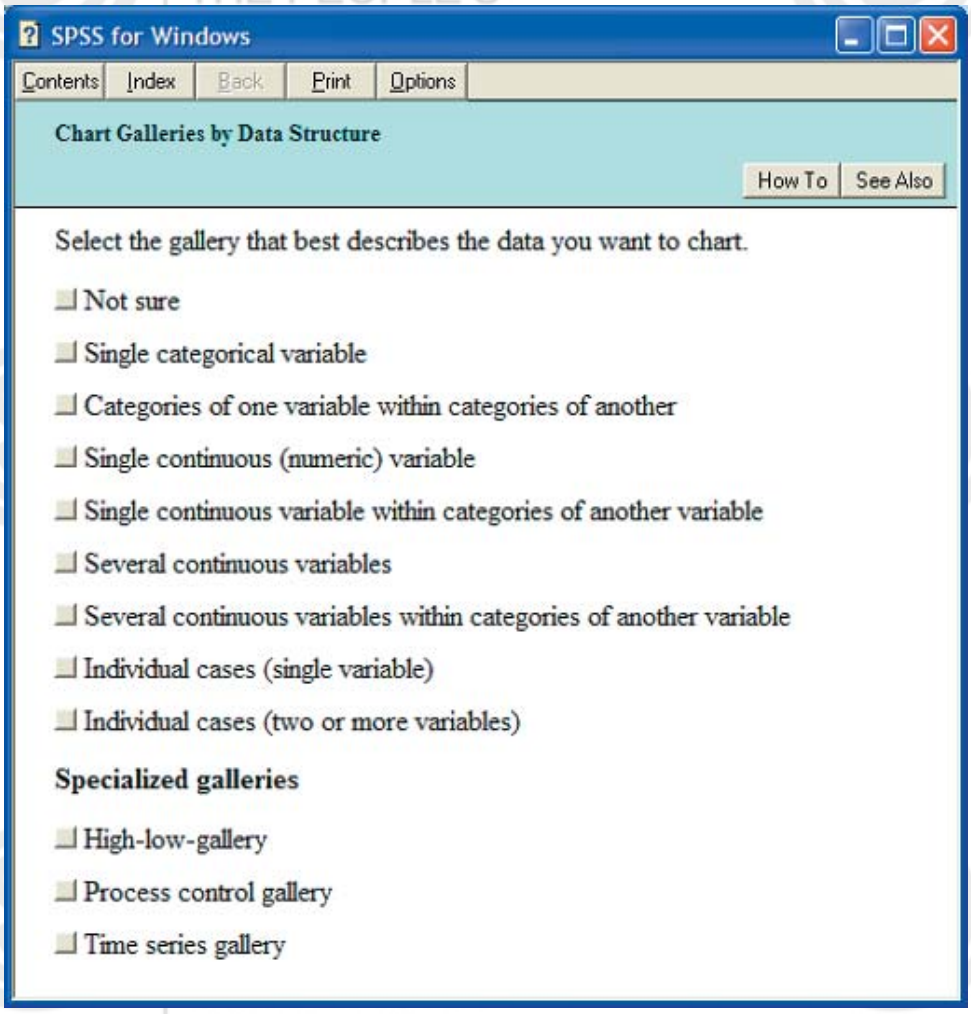

Fig. 9.9: ScreenSnapshot of Chart Galleries by Data Structure in SPSS

Now suppose you have chosen single categorical variable as the gallery that best describes your data. Then the next screen that will appear would be like the following figure. Suppose here you choose Simple Pareto Counts or Sums for Groups of Cases, then SPSS will describe what this graph does like, "Creates a bar chart summarizing categories of a single variable, sorted in descending order. A line shows the cumulative sum."

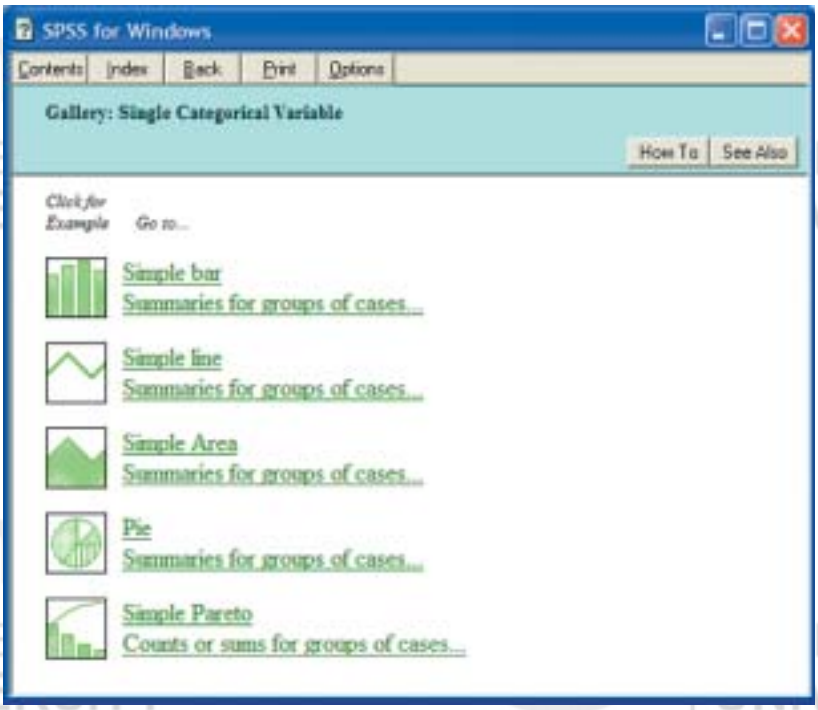

Fig. 9.10: Screen Snapshot of Gallery : Single Categorical Variable

So in this way, SPSS also acts as a mentor also. Probably, this is the reason for its success also. SPSS has a menu called "Statistics Coach", which asks questions about your data like "What do you want to do with your data?"

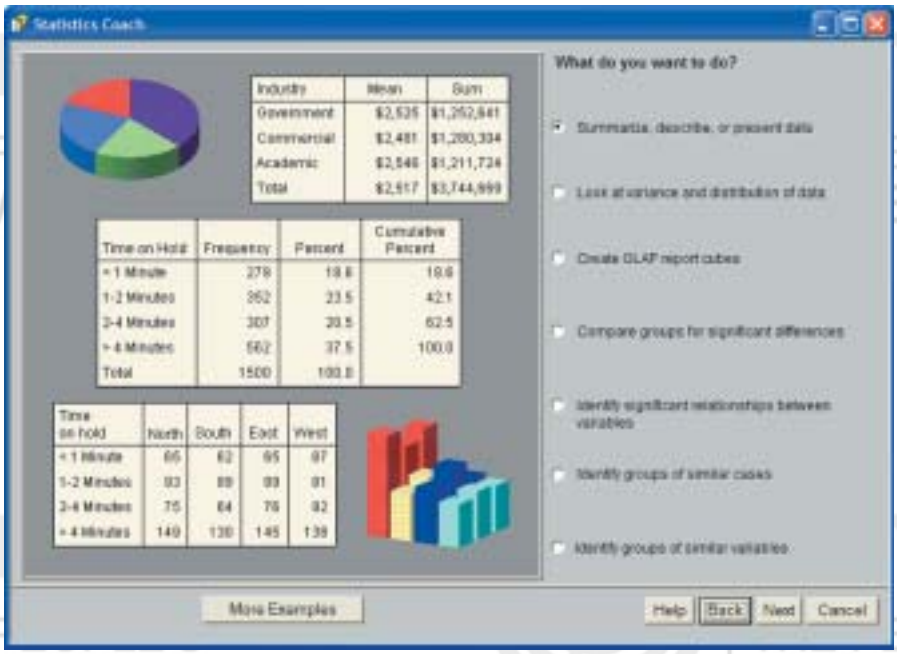

Fig. 9.11: Screen Snapshot of Gallery : Single Categorical Variable

It asks you further questions about your data in four steps and then suggests the right kind of analysis for your dataset. The output of SPSS appears as pivot table, which can be cut and pasted to Word documents, Excel worksheets and PowerPoint presentations. According to Wegman and Solka (2005), "SPSS supports numerous add-on modules including one for regression, advanced models, classification trees, table creation, exact tests, categorical analysis, trend analysis, conjoint analysis, missing value analysis, map-based analysis, and complex samples analysis".

Self Check Exercise

3) Enumerate the statistical analysis provisions in SPSS.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

9.4 OTHER SOFTWARES FOR STATISTICAL ANALYSIS

Modern statistics can perform very large and complex calculations with the help of computers. There are a lot of softwares available in the market. Many of them are shareware, freeware and online pages that perform statistical calculations. Many of the universities offer statistics online computational resources e.g. http://www.socr.ucla.edu/ (University of California at LA), Statlib: Data, Software and News from the Statistics Community http://lib.stat.cmu.edu/ (Carnegie Melon University) and free statistical tools on the web http://www.cbs.nl/isi/ (International Statistical Institute). The list is endless. We therefore restrict ourselves by giving brief introduction to some of the popular statistical softwares. The table below presents this:

Table 9.4: Brief Description of Some Statistical Softwares

S.No.	Software and its URL	Brief Description
1.	The SAS System http://www.sas.com/	SAS evolved in the late 1960's at North Carolina State University. It has now become a system for complete data management and analysis. SAS represents the Microsoft of the statistical software companies. The SAS website claims that their software resides at 40,000 sites worldwide including 90 percent of those companies on the Fortune 500 list. SAS has given due importance to recent "statistical-like" advances like data mining. It has integrated mathematical/statistical methodologies, database technology, and business applications in an effective manner to remain at the top of the commercial statistical software arena.
2.	BMDP http://www.statsol.ie/bmdp/bmdp.htm	BMDP originated during the 1960s as a bio-medical analysis package. It still remains a clear favorite of biomedical field. BMDP has aligned itself with a number of other currently popular statistical products including StatXact 5.0, LogXact 5.0, SOLAS, EquivTest, SigmaPlot, Meta Analysis, and SYSTAT.
3.	S-PLUS http://www.insightful.com/	S-Plus is based on the statistical analysis language S. S-PLUS provides the user with a fully extensible environment by supporting the user to develop their own functions using the S-PLUS language. S-PLUS contains over 4,200 data analysis functions, which implement modern and robust statistical procedures. S-PLUS has superb graphical capabilities. It receives a strong support from the academic and commercial sectors.

S.No.	Software and its URL	Brief Description
4.	MINITAB http://www.minitab.com/	MINITAB combines an array of statistical methods, graphics tools, and project organization features in a user-friendly package. It combines the user-friendliness of MS Excel with the ability to perform complex statistical analysis. Thousands of successful companies worldwide, including GE, 3M, Ford Motor Company, and the leading Six Sigma consultants, use MINITAB to make data-driven decisions and achieve world-class quality.
5.	GLIM http://www.nag.co.uk/stats/GDGE_soft.asp	Generalized Linear Interactive Modeling package (GLIM) is a flexible, interactive statistical analysis program developed by the Royal Statistical Society. GLIM is not just a modeling package; it also contains many of the standard statistical procedures and has high-resolution graphics. Professionals, scientists and statisticians worldwide respect GLIM.

Source : **Wegman & Solka (2005) and websites of the softwares.**

So you have seen that there are many statistical packages that provide the state of the art facilities for performing statistical calculations. All these software are extremely good and it is for the user to work on the one on which he feel most comfortable. There is a competition among these software for providing enhanced statistical functions, enhanced user-friendliness, better graphics and sound technical support. Also there is trend to move towards "statistics-like" disciplines e.g. data mining.

9.5 SUMMARY

The science, which deals with numbers, is statistics. It crunches the numbers and organises them in a meaningful way so that information is generated. Computer can help immensely in the statistical analysis. There exists numerous statistical tools available and the need is to identify their actual usage. Most of the Statistical methods are based on the quantitative data. One can find different statistical packages for applications to different disciplines. In this Unit you have read about two such packages MS Excel and SPSS. The Unit has discussed some of their applications in details. You have also gone through a brief introduction of some of the popular statistical softwares. This Unit was intended to make you familiar with the basic statistical functions that can be performed with the help of computer and to arouse your interest in the beautiful and huge world of statistical computing.

9.6 ANSWERS TO SELF CHECK EXERCISES

1) A statistical package is defined as the software used to collect, organise, interpret and present numerical information. The need of a statistical package arises due to the complexity of calculations involved therein for analysis and inference. It helps to bring accuracy in results.

2) The data analysis tools in MS-Excel are : ANOVA, corrilation covariance, Descriptive Statistics, Expowertial Smoothing, F-Test, Moving Average, Regression Analysis, sampling Analysis, t-Test and z-Test.

3) The statistical analysis tools in SPSS are : Report, Descriptive Statistics, Compare Means, General Linear Model, Correlate, Regression, Classify, Data Reduction, Scale, Non-Pavansetric Tests and Multiple Response.

9.7 KEYWORDS

Statistics : It is a broad mathematical discipline which studies ways to collect, summarize and draw conclusions from data. In other words, it is used as a "collection of numerical facts or data".

Statistical Package : It is software for the collection, organization, interpretation, and presentation of numerical information.

Nominal Scale : Nominal measurement consists of assigning items to groups or categories.

Ordinal scale : Measurements with ordinal scales are ordered in the sense that higher numbers represent higher values.

Cardinal Scale : On interval measurement scales, one unit on the scale represents the same magnitude on the trait or characteristic being measured across the whole range of the scale.

Ratio scale : Ratio scales are like interval scales except they have true zero points. Ratio variables are very similar to interval variables.

Spreadsheet : It is a big worksheet (in many rows and columns). This worksheet can be used for data entry and for performing calculations by click of buttons.

9.8 REFERENCES AND FURTHER READING

Box, George E.P., William G. Hunter, and J. Stuart Hunter (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley and Sons.

Johnson R.A., Wichern D.W. (2003). *Business Statistics: Decision making with Data*, John Wiley & Sons (Asia) Pvt. Ltd. : Singapore, 2003.

Levin R.I., Rubin, D.S. (2001). *Statistics for Management*. Prentice Hall of India Private Limited : New Delhi.

Microsoft Office Excel 2003. *Help Documentation.*

Pottel Hans. *Statistical flaws in Excel.* Innogenetics NV, Technologiepark 6, 9052 Zwijnaarde, Belgium, http://www.mis.coventry.ac.uk/~nhunt/pottel.pdf. Downloaded 08/11/05

Ripley Brian D.(2004). *Statistical Methods Need Software.* Netherlands Statistical Society, 27 April 2004. http://www.stats.ox.ac.uk/pub/bdr/NethStatSoc.pdf Downloaded 08/11/05

Sarma K.V.S. (2001). *Statistics made simple- Do it yourself on PC*. Prentice-Hall of India: New Delhi.

SPSS for Windows Release 11.0. *Help Documentation.*

Wegman Edward J., Solka Jeffrey L. (2005). *Statistical Software for Today and Tomorrow*. Center for Computational Statistics, George Mason University Fairfax, VA 22030. Downloaded 16/11/05 http://www.galaxy.gmu.edu/papers/guide.pdf