

USE OF STATISTICAL METHODS IN AGRICULTURE AND ALLIED FIELDS

Urmil Verma

CCS Haryana Agricultural University
India

Statistical Science is concerned with the twin aspect of theory of design of experiments and sample surveys and drawing valid inferences there from using various statistical techniques/methods. The art of drawing valid conclusions depends on how the data have been collected and analysed. Depending upon the objective of the study, one has to choose an appropriate statistical procedure to test the hypothesis. When the number of observations is large or when the researcher is interested in multifarious aspects or some time series study, such calculations are very tedious and time consuming on a desk calculator. In this context, it is essential that the manpower engaged in teaching and research is to be trained in the applications of various statistical techniques/methods through the use of computer. An attempt has been made to cover computer aided analysis (using various statistical packages) related to Descriptive Statistics, Test of Significance, Design and Analysis of Experiment, Non parametric method, Forecasting through time-series models and some Financial analysis etc. A healthy group discussion (through practical exercises) can also be held on most commonly used statistical techniques. Computing platform will involve both the environment i.e. DOS as well as Windows 2000.

INTRODUCTION

In the present scenario, information is a resource, and like any other resource, its proper, efficient and effective use is the key to success. Information is extracted from data after its proper processing. We are gathering a lot of data every day by putting in considerable efforts/money. Next is the question of its utilization/processing. The huge amount can't be processed timely and efficiently through manual methods. Computers can help in this regard, perfectly. In almost all disciplines like education, agriculture, medicine, business etc. data are pivoted in decision making. Hence for a good study, proper collection, analysis and interpretation of the result, it is essential to prepare a cogent report. Raw data handling, its tabulation and preliminary analysis is the most time consuming, cumbersome and boring task in research. Normally, while doing it manually on a square sheet, one has to be careful in writing the data, as one wrong entry in data may give disastrous results. Use of an Electronic Spreadsheet makes this whole, easier and enjoyable task.

An electronic spreadsheet is like a conventional paper spread sheet. It is a table consisting of columns and rows of entries. These rows and columns can be easily manipulated and set up so that calculations can be performed on them. A great advantage to electronic spread sheet program is that if one makes a change in one place, all other related figures are automatically recalculated. The automatic recalculation makes it a simple matter to correct mistakes quickly, update values and try out any number of 'what if' hypothesis without having to perform repetitive manual calculations. Excel, the most popular spreadsheet package, is a powerful yet flexible tool that can be applied to a wide range of common problems.

STATISTICAL ANALYSIS THROUGH MS- EXCEL STATISTICAL FUNCTIONS

A bulk of statistical functions related to Descriptive Statistics as well as Mathematical, Financial and engineering worksheet functions can be made applicable through MS- Excel functions icon. A list of available worksheet functions can be seen by clicking Edit *Formula*, on formula bar and then click the down arrow in *Insert Function*.

STATISTICAL ANALYSIS THROUGH ANALYSIS TOOLPAK

Microsoft Excel provides a set of data analysis tool called the *Analysis ToolPak* that one can use to save steps when one develops complex statistical or engineering analyses. One should 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 table. To view a list of available analysis tools, click *Data Analysis* on the *Tools menu*. If the *Data Analysis* command is not on the *Tools menu*, run the *Set up* program to install the *Analysis ToolPak*. After installing the *Analysis ToolPak*, one must select it in the *Add-in Manager*.

TO INSTALL & USE THE ANALYSIS TOOLPAK

Before using an analysis tool, the data to be analyzed must be arranged in columns or rows on the worksheet. This is the *input range*. If the *Data Analysis* command is not on the *Tools menu*, one need to install the *Analysis ToolPak* in Microsoft Excel. The steps to be followed are as:

- i) On the *Tools menu*, click *Add-Ins*. If *Analysis ToolPak* is not listed in the *Add-Ins dialog box*, click *Browse* and locate the drive, folder name and file name for the *Analysis ToolPak add-in*, *Analys32.xll* _ usually located in the *Library\Analysis folder* _ or run the *Set up* program if it isn't installed.
- ii) On *Tool menu*, click *Data Analysis*
- iii) Select the *Analysis ToolPak* check box and click the tool to be used for the purpose.
- iv) Enter the *input range* and the *output range* and then select the option of one's own need.

Several analyses tools available are Covariance, Correlation, Regression, Test for Statistical Significance i.e. t-test (Two sample assuming equal variance analysis as well as unequal variance analysis, t-test (Paired, two samples for means analysis, An analysis of Variance (ANOVA) (for single factor, two-factor with replication and without replication), Sampling analysis and Descriptive Statistics etc. Descriptive Statistics analysis tool generates a report of univariate statistics for data in the input range, providing information about the central tendency and dispersion of the data. The basic thing common to all the analysis are:

Input Range. Enter the *cell reference* for the range of data to be analyzed. The term *Grouped By* indicate whether the data in the input range is arranged in rows or in columns, click *Rows* or *Column*.

Output Range. Enter the reference for the upper left cell of the output table. Microsoft Excel writes a two-column table of statistics for each column or rows in the input range, depending on the *Grouped By* option selected.

New Workbook. Click to create a new workbook and paste the results on a new worksheet in the new workbook.

Some general packages like SPSS, SAS are used for analysis of data obtainable in such situations individually, that is, treating each as a separate case. But to use such general package, it is necessary to be well conversant with the methods of analysis appropriate in each individual case, otherwise, instructions for using the package can not be formulated properly. A better, easy and straight-forward way for providing computer solutions for all such data is to evolve suitable customized general program capable of handling data from any such designs/situations, with at least interactions with the computer. There are many small statistical packages like *SPSS*, *CPCS*, *M-Stat*, *Spar1*, *OPSTAT* etc., simply of single floppy program, that can be used for the same purpose. A few tips on SPSS and CPCS are explained below:

STATISTICAL ANALYSIS USING SPSS IN DOS ENVIRONMENT

The SPSS text Editor

SPSS/PC comes with its own editor, called REVIEW, which is built into the SPSS working environment. One can use some different editor or word processor, instead of REVIEW. Just be sure to the file as ASCII or text files, with no special formatting codes. To start up REVIEW, enter the following REVIEW test.dat. Review will be ready to enter the data in the test.dat file to store the data.. There are a number of function keys labeled F₁, F₂ and so forth, on the keyboard. One can use REVIEW help by pressing F₁. The commands are readily available from the mini-menus, and the help displayed.

Starting up the system

It is presumed that the SPSS program is loaded on the hard disk. To start it, type *student* and press *enter*; C: student; the screen will be displayed as under; for help, type *HELP*.

SPSS/PC: Quitting, to get out of SPSS/PC, enter the command *Finish* just to turn to DOS *Identification and analysis of data*. To do anything useful with the data, one need at least two SPSS commands, one to identify the data and another to analyze it.

Identification of data file. The first thing one has to do on the DATA LIST command is to identify the data file one is going to use. Type the word *DATA LIST*. Next, enter the keyword FILE, an equal sign and the name of the file enclosed in apostrophes. After this, the variables to be analyzed are defined in the following manner. *Data list file = 'test.dat' / id 1-4 bmonth 6-7 bday 8-9 Sex 12*. i.e. id is identification number appearing in column number 1 to 4 and so on. Save the defined file by giving the save outfile command. Save outfile = 'test.sys'. 'test.sys' is called a system file.

One can proceed further for analysis of the data. After quitting the SPSS, the system file may be recalled using the *Get File* command. Get file = 'test.sys'. Every command should end with a period '.'.

Analysis of data. Analysis in relation to descriptive statistics, test of Significance, correlation, regression, design and analysis of experiment etc. can be had with the help of this package. The way of writing the commands for some important type of analysis commonly used by the researcher is given as below:

Correlation

CORRELATION varlist WITH varlist / options = n / statistics = n.

Crosstabulation

CROSSTABS varlist BY varlist / option = n / statistics = n.

Frequencies

FREQUENCIES varlist/BARCHART/HISTOGRAM/Statistics=All.

One way analysis of Variance

ONEWAY varlist BY varname (min. max)/CONTRAST/ Ranges=LSD DUNCAN/stat = n.

Regression

REGRESSION varlist / dependent = y / method = enter forward / statistics = dflt.

STATISTICAL ANALYSIS USING CPCS IN DOS ENVIRONMENT

The program computes necessary statistics concerning Design and Analysis of Experiment, with equal or unequal number of replications. Analysis of the data pooled over environments is also provided. Data can be analysed using arcsin, square root, natural logarithms or kilogram per plot to quintals per hectare transformation. If desired, the program can also be used for missing data. The package provides analysis on

Completely Randomized Design (CRD)

Randomized Block Design (RBD)

Split-plot Design (SPD)

Factorial Experiment in CRD, RBD, SPD

Analysis of Covariance in CRD or RBD

t-test, paired t-test etc.

INSTALLATION

To install this package create a directory and copy all the files from source diskette to this directory. A user can execute the program by typing CPCS at the DOS prompt and press <enter>, a main menu will appear on the screen with a number of options and the user has to enter the option according to his or her own need. Before executing this software one has to ensure that the data file has been created in a specified manner.

CREATION OF DATA FILES

For creating a data file, the MS-DOS editor can be invoked by typing 'edit<filename>' at the command prompt. This file can be saved by 'Alt+F+S' and one can exit MS-DOS Editor by pressing 'Alt+F+X' keys.

FORMAT OF DATA INPUT

- i) Identification of the data set; first line of data is provided for identification of the data set of which first 80 columns will be reproduced as such in the output.
- ii) Parameters of the data set; Values of the following parameters are to be supplied in second line of the data set in the given order with at least one blank among them.
 NR NT NL(J) JT PS
 Where,
 NR = number of replications
 NT = number of factors
 NL(J) = number of levels of Jth factor
 (J = 1,2,3...NF)
 JT = 0, when transformation is not required
 = 1, for transformation from kg/plot to q/ha
 = 2, for arc sin transformation
 = 3, for square root (n+1) transformation
 = 4, for natural logarithmic transformation
 PS = area of plot in square meter when JT =1
 = 0, otherwise
- iii) From third line onwards, data of the set are to be supplied for all treatment combinations starting with first level of each factor and varying all levels of the last factor, then the last but one factor and so-on, giving serial number and the values of all the replications for each treatment combination in one line.
- iv) Repeat steps 1 to 3 for more set of data

SUMMARY AND CONCLUSIONS

Statistics has its universal application as an essential tool in various fields of research, for exp., use of different statistical techniques for the critical analysis and comparison of the experimental data pertaining to the multifarious aspect on basic and applied research through the use of different softwares. The packages are simple and easily accessible even for ordinary user of the computer and do not require any specific know how. The first step is to create a data file in the MS-DOS Editor. The tables which are used for the analysis are prepared according to the standard methods, which are being used generally in the routine statistical analysis computation. The data is entered in the input data file keeping a single space between the observations in all the techniques mentioned in a particular package in free format. The second step is to approach the program by typing its name at the command prompt. The option will appear. Enter the serial number of the option which follows some required statements pertaining to the technique for which, one has opted. Supply this information. The results will be stored in the output file. Type 'Edit<outputfile name>' and press enter. The screen will display all the results of the experiment. The results can be printed by pressing Alt+F+P keys.