

'STATTECH'—A COMPUTER PROGRAM PACKAGE FOR STATISTICS*

By DEBASHIS ROY and AMITAVA DATTA

Indian Statistical Institute

STATTECH, a program package for statistical data analysis together with management and quality control charts preparation has been developed and announced by SOFTWARE ULTRATECH SYSTEMS (P) LTD, a Madras based software production company under the direction of Dr. M. N. Murty, an applied statistician of international reputation. It runs on IBM compatible Personal Computers (PCs) with minimal configuration under MS-DOS and/or PC-DOS Operating Systems. The current released version is 1.1.

STATTECH, consisting of four parts, is intended for people with some background of statistics and usage of computers. Each part is provided on a separate floppy with a manual describing operations and restrictions.

STATTECH-1 covers Statistical Analysis which provides for—Multi-way Tables ; Descriptive Statistics ; Linear, Non Linear, Multiple Linear and Polynomial Regression Analysis ; Analysis of Variance ; Time Series Analysis ; Index Numbers ; Rank Correlation and Random Sampling.

STATTECH-2 covers Graphs and Charts which provides for—Pie Charts ; Bar Charts ; Column Charts ; Line Graphs ; Scatter Diagram ; Frequency Distribution Graphs and Charts ; Time Sequence Charts and Pareto Diagrams.

STATTECH-3 covers Quality Control Tools which provides for—Histograms ; Scatter Diagrams ; Pareto Diagrams ; Control Charts and AQL Inspection Plans.

STATTECH-4 covers Experiments and Inference which covers—Analysis of Variance ; Basic Designs ; Nested Designs ; Factorial Experiments ; Special Designs ; Analysis of Covariance ; Large Sample Tests ; Small Sample Tests and Non-parametric Tests.

**Editorial Note :* The Editors are grateful to Professor J. Roy for handling this review as a guest editor. *Sankhyā* intends to publish such reviews in future.

AMS (1980) subject classification : 62-04, 62-07.

Key words and phrases : Statistical data analysis, well behaved data, Ms-Dos, PC-Dos, program package, Menu driven, prompt message, command mode.

Though STATTECH is versatile and complex, a simplified appearance has been provided for users. The package is menu driven within each component, and the user interacts with the programs in simple query-response and/or option selection. And unlike many of the commercial packages, it is reasonably well documented.

A series of tests were carried out at the Computer and Statistical Service Centre of the Indian Statistical Institute to check the performance of this package. Though the testing cannot be claimed to be exhaustive, as it is virtually impossible to carry out an exhaustive test for such a versatile package with so many optional paths within finite time and resources, quite a few peculiarities were observed. Our observations are presented hereafter.

Most of the programs run properly with well behaved data. The programs for Analysis of Variance, Multiple Linear Regression, Non-linear Regression, displaying Scatter Diagram only, etc. failed to execute properly during the tests.

The programs, however, fail with singular or near-singular data. In most such situations, the programs go to Command mode of BASIC with the message "divide by zero at ...". The results are also erroneous where programs permit the user to select multiple variates one after another from a multivariate data—the moments of one variate shown by the program may be that of the previously selected variable.

Under STATTECH, Program selection and execution is menu driven and data entry is through query and response, which make the package easy to handle. Error messages displayed during program execution are, however, those produced by BASIC compiler and/or interpreter. Thus, if an user is not familiar with BASIC programming, he/she may not be able to guess what is wrong, as the message may be quite confusing. After displaying the error message like "subscript out of range at ..." or "divide by zero at ...", program under STATTECH-1 returns to DOS mode, while programs under STATTECH-2, STATTECH-2, STATTECH-3 or STATTECH-4 goes to Command mode of BASIC.

The prompt messages that appear on the screen while executing the package at times do not match with the operations manual. Some of the programs also do not work properly if the user responses in the format described in the operations manual. In some cases, queries placed on screen are not described in the operations manual. Valid options are not

supplied within [] in a few situation. A list of such situations is provided later in this document. These may create problems for users who are not experts in both statistics and programming, but would like to use the package.

Long prompts, if properly arranged in multiple lines, would appear more readable than abrupt termination and continuation which is the case at present.

User is expected to respond with either YES or NO to many a query prompted by STATTECH. Responses in small letters, y/n, in place of Y/N, are either not accepted or some default is taken by the programs. As these defaults are neither mentioned in the documents, nor in the prompts, an user may often traverse a non-opted path.

The programs do not check input data. As a result, if erroneous data like negative frequencies and/or negative counts, upper bounds less than lower bound of class limits, etc. are provided, and user does not check these properly, peculiar results, graphs, shadings, etc. are observed. Though the operations manual claims that “—” would be replaced by 9999.99 or nearby value, this does not happen.

Graphs and Charts under STATTECH-2 and STATTECH-3, if directly dumped onto a printer by pressing <SHIFT> and <PRTSC> keys together as claimed in the operations manual, do not get printed properly. These are printed properly only if the DOS utility GRAPHICS is loaded onto memory before STATTECH is initiated.

Data created through one program may or may not be accessed by another program. This inconveniences the user if he/she wants to analyze and/or compute different statistics using the same data. The user may have to enter the same data more than once. If the data be voluminous, this is time consuming and error-prone.

Data correction facilities do not extend for correcting the beginning entries like Physical-Unit name, Unit name, Block type, Number of Blocks, Time-Frame etc.

A few peculiarities are observed with saving the data entered under a program. The programs terminate with or without any error message if the floppy on which data are to be saved is write-protected or not inserted.

If data is read from a file and edited through the correction procedure, program stops with “file already open” message during data processing if the modified data is not saved. A retry option along with proper checking in such situations would be very helpful.

While using the package using an IBM-compatible PC, <[2.J often appears on the left hand side of a new line on the screen when <ENTER> key is pressed. No such line appears when a PC-XT or a PC-AT is used. We have not been able to identify the reason of such anomaly.

Only Laspayre's index may be calculated under the program of Index Number computation, which may be grossly inadequate for many a user. We feel that if provisions for computation of a few more indices like Paasche's, Fisher's, Wold's, etc. and particularly indices with more general weightage procedure, like Cost of Living Index, Wholesale and Retailer Index, etc. be provided, the flexibility and usability of the package would increase tremendously.

Observations on STATTECH-1 :

After a Graph is displayed, the next prompt message appears along the X-axis description. This mars the tidy appearance of the Graph.

In the Descriptive Statistics module, if all constant values are input, the program aborts with error message "Division by zero". For ungrouped data with grouped processing, if 'Y' pressed against the prompt: "WANT TO STORE THE VALUES IN A FILE (Y/N)", it aborts with error message "Division by zero". With ungrouped processing of ungrouped data the program, after printing the charts and tables, the program loops back to displaying the prompt when the option for exiting is entered against the prompt "CHOICE OF VARIABLE" AND "END PROGRAM".

For the Multi-Way Frequency Table module, there are a lot of discrepancies between the program and the operations manual. The first paragraph of the operations manual says that for "Transformation", chapter 10 should be referred to. Chapter 10 of the operations manual, however, does not describe the Transformation rules. The second paragraph in the operations manual says that the user should input 0 as the no. of class intervals, if he/she does not require the frequency table for a variable. At the time of frequency table creation, the program, however, prints a table with all entries as zeroes. The option "Choose 1 if any correction needed, ... correction routine" in the Multi-Way Frequency Table module, does not tally between the menu displayed and the operations manual. The option number wise details also do not tally.

The Linear and Non-linear Regression module asks for Variable name, Unit of measure and Time frame for each variate, while the operations manual indicates that Time frame is to be entered only once.

In the Prediction Routine, "Type Mismatch Error" occurs after the values of the auxiliary variables for the required number of units has been keyed in.

For the Multiple Linear Regression module, when any method—either optimal or non-optimal, is selected and the number of auxiliary variables have been keyed in, error message "File Not Found" appears. The same error message appears for Polynomial Regression also. This happens even with the sample data supplied in the operations manual.

While testing the Analysis of Variance module, error message "File Not Found" is displayed at the very beginning.

In the Time Series Analysis module, the queries displayed on the screen do not tally with that described in the operations manual. Either the sequence of prompts is different or some prompts are dropped. For example, a prompt "TYPE PHYSICAL UNIT/GROUP ..." is displayed on the screen which is not described in the operations manual. The operations manual claims that the period for Time Series Analysis should be less than $N/3$, the program does not check this condition and accepts value more than $N/3$. When seasonal variation is opted for in the Time Series Analysis module, the program aborts with error message "Write Protect on Drive B". The response to the prompt "TYPE NO. OF VARIABLES, NO. OF YEARS AND PRESS <ENTER >" does not seem to correspond to what is described in the manual.

Observations on STATTECH-2 :

All the graphs displayed on screen are printed properly only if the GRAPHICS routine is loaded before STATTECH is initiated.

Pie Charts printing program do not print a circle if only one variate and one value is given as input, it prints a line.

Additional query "PHYSICAL UNIT/GROUP TO WHICH IT BELONGS ?" is displayed in the Scatter Diagram module, which is not explained in operations manual.

The Scatter Diagram is not displayed at all. Program stops with error message "Subscript out of range in 1260" after displaying the axes. This happens with all data—user created or sample provided in the manual.

Data-File created for Scatter Diagram program cannot be read in the Frequency distribution module, system hangs. In this program if input data be read from a file and modified without re-saving, the program stops with message "File already opened at 1650".

The Frequency distribution module permits repeated selection of variables. Skewness and Kurtosis values are not initialised in between. If a variate is singular (all equal values), skewness and kurtosis of the earlier selected variate is at times displayed, instead of zero. If Class-intervals are not provided in this module, the program always prepares 3 class intervals, even when there is only one value repeated multiple times.

Observations on STATTECH-3 :

All the graphs displayed on screen are printed properly only if the GRAPHICS routine is loaded before STATTECH is initiated.

In the Histogram module, even when Unit Data is opted for, the prompt "sub-group no." is displayed for data input, which creates confusion. The prompts for No. of Class Intervals, Lower limit of 1st Class Interval, and Length of Interval are displayed for all the variates, after user has keyed in the particular variate he/she wants the histogram for. The order of the prompts and the operations manual both give a wrong impression on this point.

The equation of the regression line is always displayed in the Regression module, even when the line is not opted for. In such a situation, the line is not shown, but the equation is. And the equation of the regression line does not match with the line drawn.

Observations on STATTECH-4 :

In "ANALYSIS OF VARIANCE" menu, item no. 3 says "2-Way classification with multiple but equal values", whereas in introductory part of the operations manual, it says "2-Way classification with multiple but equal no. of observations per cell". With the demonstration data provided for this module in the operations manual, the program terminates with error message "Illegal function call" after classification no. and group no. are both entered as 2.

The prompt "TYPE NO. OF OBSERVATIONS PER CELL AND PRESS <ENTER >" is repeatedly displayed in the Basic Design module, though the number of observations should be equal for all the cells. In the same module, at step number 10, the operations manual says "Skip next three prompts", whereas the next five prompts should be skipped for obtaining proper results. Step no. 26 of this program is not prompted at all.

In Special Design module the options indicated within [] are not consistent with that described in the tutorial example. The printout format obtained for Analysis of Variance in step no. 26 is not proper. And the error

message "Out of String Space" appears while trying to display the ANOVA table.

Files created in STATTECH-1 cannot be accessed properly in Large Sample Test modules of STATTECH-4.

In conclusion, we declare STATTECH to be a versatile package. It covers almost all the aspects of statistical analysis that non-statisticians (and statisticians too!) usually carry out in day-to-day operations. We hope that the problems observed with the current release would be rectified in the subsequent release.

Paper received : May, 1989.