A REVIEW OF MICRO-COMPUTER STATISTICAL PACKAGES FOR USE BY TRANSPORT PLANNERS

WILLIAM YOUNG
Lecturer
Department of Civil Engineering
Monash University

M.A.P. TAYLOR
Senior Research Scientist
Division of Building Research
CSIRO

ABSTRACT: *The present level of sophistication in transport planning stems largely from developments in the computer industry. The current trend in the development of micro-computers will also influence transport planners. Their ability to solve problems, prepare reports and work more efficiently should increase. One aspect of considerable interest to transport planners is that of statistical analysis of data. Many statistical computer programs have been developed for large computer installations and it is therefore to be expected that similar packages will be developed for micro-computers. This paper reviews the statistical needs of transport planners and investigates the role statistical packages, developed for micro-computers, play in meeting these needs.*

A REVIEW OF MICROCOMPUTER STATISTICAL PACKAGES

INTRODUCTION

There is a trend in the computer industry towards computers becoming cheaper and smaller in physical size, whilst maintaining a considerable level of computing power (Richardson, Kok and Vandebona, 1983). This trend is resulting in more and more transport planners obtaining their own personal computer systems. Their ability to solve problems, prepare reports and work more efficiently is therefore increasing. Alongside this trend is a growth in the development of software to suit these personal computers. One type of software of considerable interest to transport planners is that related to data interrogation. This requires both data management and statistical routines. This paper focuses on the latter of these types: statistical packages.

Because of the rapid developments in the computer field it is inevitable that any review of these packages will probably be out of date by the time it is published. This review does, however, provide an indication of the range of packages available, the directions in which the packages are heading and a number of aspects that prospective users should be aware of. Further, although the authors have attempted to be as comprehensive and accurate as possible, they have undoubtedly missed some aspects or they have misinterpreted some of the information available to them. It would therefore be appreciated if the readers aware of any such deficiencies inform either of the authors.

The paper is divided into three parts. The first reviews the statistical needs of transport planners and considers the routines available in a number of the packages for mainframe computers. The second section provides a list of packages available for micro-computers along with the statistical procedures they contain. A number of comments on these packages are made. The third section considers a commonly used micro-computer statistical package (MICROSTAT) in some detail.

STATISTICAL NEEDS OF TRANSPORT PLANNERS

Introduction

Transport planning is a complex process requiring input from a multidisciplinary team. Its practitioners may be employed in tasks that range from gap acceptance studies to developing complex behavioural or prescriptive models of transport systems. The statistical needs of each planner may therefore be quite specific while the overall needs of a group of transport planners may be very diverse. To obtain an indication of the breadth of these needs this section presents a review of some of the techniques discussed in the transport planning literature. Emphasis was given to techniques discussed in previous Transport Research Forums. The availability of mainframe computer packages to meet these needs is also discussed.

Statistical Needs

Statistical procedures may be divided into three general groups. These are data investigation, hypothesis testing and relationship determination. The data investigation routines provide the procedures for initial data scanning and are therefore very important parts of any statistical package. The hypothesis testing routines enable the comparison of data sets while the relationship determination routines allow the parameters of mathematical models to be estimated. Routines found in each group are outlined below.

The data investigation routines provide an initial look at the data. They can therefore influence the development of the approaches used in later stages of analysis. The statistical procedures found in this group start with

(a) measures of central tendency like means, medians and modes.

(b) measures of variation like standard deviations and coefficients of variation

(c) measures of asymmetry such as skewness and kurtosis.

Following this group of procedures are methods for more detailed examination of the distribution of the data. Routines for preparing frequency tables and diagrams provides the first level of analysis. The second level investigates distributions over two or more variables, and requires routines to prepare and present scatterplots and cross tabulations (Taylor, 1978).

Hypothesis testing routines take statistical analysis a step further. However, before discussing these procedures a distinction between the various types of data needs to be drawn. When data are collected, the process of assigning a score or value to the observation constitutes the process of measurement. The rules defining the assignment of the value determine the level of measurement. There are four levels of measurement that are commonly used to describe data (Carterette and Friedman, 1974). These are nominal, ordinal, interval and ratio. The nominal level is the lowest level since no value is associated with the data. Rather each value is placed in a distinct category (e.g. the numbers placed on buses to distinguish their route). The ordinal level ranks data in some order (e.g. the order buses arrive at a terminal). Interval scales have the property that the distance between each category is defined in terms of equal units (e.g. time of arrival of buses at a terminal). The final level, ratio scales, have all the properties of the above scales as well as having a definite zero point (e.g. the engine capacity of buses). The properties of these scale types are summarised in Table 1.

Hypothesis testing with each level of data requires a different approach. The nominal level can use contingency table analysis and chi-squared tests to see if there are differences in the data describing each nominal group. Non-parametric tests can be used to test hypotheses with respect to ordinal data. Interval and ratio scales possibly offer the greatest array of statistical procedures for hypothesis testing. Tests have been developed to compare the means of variables whilst relationships between variables can be tested using correlation analysis (Abelson and Baker, 1982) and analysis of variance (ANOVA).

Another element of hypothesis testing relates to goodness-of-fit. Procedures for testing the difference between theoretical distributions and observed data are available for all four levels and are an important part of the statistical needs of the transport planner.

Table 1. Summary of Scale Types

Scale Type	Central Tendency	Statistics Variability	Individual Position	Permissible Uses	Permissible Transformations
RATIO	Geometric Mean	Coefficient of Variation	Absolute Score	Find ratios between	Multiplication and division
INTERVAL	Arithmetic Mean	Variance, Standard Deviation	Relative Score	Find differences between	Addition and subtraction
ORDINAL	Median	Range	Rank Percentile	Establish rank order	Any that preserve order
NOMINAL	Mode	Number of categories	Belonging to category	Identify and classify	Substitution within category

Note: The higher level scales subsume all the features of the lower level scales.

The third set of statistical procedures relate to the determination of relationships between variables. In transport planning numerous procedures have been used to investigate these relationships. Some common procedures are:

(a) *Multiple Linear Regression*, a general statistical technique used to develop relationships between a dependent variable and a set of independent or predictive variables (Faulkner and Nelson, 1983; Bradley and Holsman, 1983). The technique assumes that there is a linear relation between the dependent and independent variables. It generally uses interval and ratio data but nominal and ordinal data can be included through dummy variables.

(b) *Stepwise Regression*, a procedure for entering independent variables one at a time into a multiple linear regression model according to some pre-established criteria (Young and Richardson, 1978).

(c) *Discriminant Analysis*, uses a linear combination of independent variables to classify cases into nominal categories. This procedure has been used to study choice of mode and other transport decisions (Stopher and Meyburg, 1974).

(d) *Logit Analysis*, like discriminant analysis, logit analysis uses a combination of independent variables to classify cases into nominal categories. Unlike discriminant analysis it uses a sigmoid relationship to separate the choice of nominal categories. This procedure is widely used in contemporary transport planning (Young, 1979; Brown, 1982; Hensher and Manefield, 1982).

(e) *Probit Analysis* is another procedure for studying transport related decisions (Daganzo, 1980). It is based on the multinormal distribution and represents a more general procedure than logit analysis.

(f) *Log-linear Analysis* permits the study of multiway contingency tables and the interactions between variables (Dumble 1979).

(g) *Time Series Analysis*, uses statistical procedures to investigate changes in particular variables with time (Aplin, 1983).

These seven statistical techniques represent methods of examining, explaining and predicting relationships between independent and dependent variables. The following techniques attempt to find the underlying dimensions of a set of data describing a large number of variables. The techniques include:

(a) *Principal Component Analysis* which uses correlation between interval or ratio level variables as a basis for grouping like variables.

(b) *Factor Analysis* goes one step further than principle component analysis since it manipulates the principle components to enable easier interpretation of the underlying dimensions (Young and Ogden, 1983).

(c) *Canonical Correlation* can be seen as a combination of linear regression and factor analysis. It has the data reduction capabilities of factor analysis, but, having required the user to divide the variables into two sets, also assesses the relationship between the two sets of factors.

(d) *Multidimensional Scaling* is like factor analysis, in that it groups together similar attributes. Its main difference relates to the level of the input data. Multidimensional scaling uses measures of similarity like paired comparisons and rank ordering (Nicolaidis, 1975).

Mainframe statistical packages

The procedures outlined in the above discussion are listed in Table 2. Alongside this list are the procedures available in three statistical packages developed for mainframe computer systems. The three packages are the General Statistical Package (GENSTAT, 1973) the International Mathematical and Statistical Library (IMSL, 1982) and the Statistical Package for Social Sciences (SPSS, 1972). Table 2 shows that each package provides a large array of statistical procedures. From the transport planning view the procedures that are overlooked in at least two packages are Multidimensional Scaling and Distribution fitting. Multinomial Logit and Probit Analysis are not available in any of the statistical packages. Special programs are usually used for these tasks.

A further difference in the three packages is their presentation to the user. SPSS is possibly the most well known since it has been prepared so that it can be used by people without a detailed knowledge of computer programming or statistics. Its user manual first discusses the statistical procedures in a simple and concise fashion then outlines the computer command statements. The main criticism of the package relates to the misuse by people without appropriate knowledge of statistics. GENSTAT is a statistical package

Table 2 Statistical Procedures in Mainframe Computer Statistical Packages

STATISTICAL PROCEDURES	STATISTICAL PACKAGE FOR SOCIAL SCIENCE (SPSS)	INTERNATIONAL MATHEMATICAL AND STATISTICAL LIBRARY (IMSL)	GENERAL STATISTICAL PACKAGE (GENSTAT)
Data Investigation			
Descriptive Statistics	x	x	x
Frequency Distributions	x	x	x
Scatterplots	x	x	x
Cross Tabulations	x	x	x
Hypothesis Testing			
Chi Squared	x	x	x
Non Parametric	x	x	
Difference in Means	x	x	x
Correlation	x	x	x
Analysis of Variance	x	x	x
Distribution Fitting			x
Relationship Determination			
Multiple Linear Regression	x	x	x
Stepwise Regression	x	x	x
Discriminant Analysis	x	x	x
Logit Analysis			Binary
Probit Analysis			Binary
Log-linear Analysis		x	x
Time Series Analysis		x	x
Data Reduction			
Principal Component Analysis	x	x	x
Factor Analysis	x	x	x
Canonical Correlation	x	x	
Multi-dimensional Scaling			x

used by many statisticians. Like SPSS it has its own control language, however, the language is more concise. The package offers considerable flexibility in the presentation of the output. IMSL is a library of statistical procedures that can be accessed by control programs. The user must therefore have both a good knowledge of computer programming and statistics before he can use the system.

In concluding this section it should be pointed out that none of these computer packages are interactive. Initial data investigation can be time consuming since the user is restricted to batch input only. Once the analysis has been completed the user must obtain a copy of the results, synthesize them then investigate the next aspect. Micro-computers provide a considerable improvement in this phase of data analysis.

GENERAL REVIEW OF MICRO-COMPUTER STATISTICAL PROCEDURES

The previous section discussed the statistical needs of transport planners and has shown how they are met by some mainframe statistical packages. This section discusses the availability of statistical packages for micro-computers and the procedures they offer.

Appendix A presents a list of twenty seven statistical packages that have been developed for various micro-computers. Although this list may not include all statistical packages developed for micro-computers it does provide an indication of the capabilities of the readily available packages. It can be seen that most packages provide procedures to determine the descriptive statistics, frequency distributions, correlation between variables and multilinear regression coefficients. Few packages provide procedures for distribution fitting, stepwise regression, discriminant analysis, logit analysis, probit analysis, log-linear modelling, principal components analysis, factor analysis, canonical correlation and multidimensional scaling.

The most comprehensive packages appear to be ABSTAT, INTERSTAT, MASS, MICROSTAT, SAM, STATPAC and STATPAK, all providing at least nine statistical procedures. Two packages that appear to be developing into very useful packages for transport planners are MATHSTAT and MDA. MATHSTAT aims to incorporate stepwise regression, logit analysis, probit analysis, factor analysis, canonical correlation and multidimensional scaling by early 1984. MDA, developed by a prominent transport researcher from Massechusetts Institute of Technology, already includes logit analysis and plan to incorporate analysis of variance and factor analysis by early 1984. However, the purchaser of any statistical package should be wary of purchasing it in the expectation that certain procedures will be incorporated in the future.

A package that appears to provide a useful extension to the developments in mainframe systems is SL-Micro. This system is a reduced version of the SPSS system. It therefore has the considerable advantages associated with this package: excellent documentation, no need for detailed programming knowledge and consistent format. One possible limitation is its lack of interactive capabilities. The program is accessed through a set of commands arranged in a sequential fashion. This set of commands is submitted as a job and the appropriate calculations carried out. The output is sent to a file which can be accessed after the job has been completed.

As well as the statistical procedures available in the packages a number of other aspects may be of interest. Appendix B presents information on the supplier's address, computer language used to develop the program, the microcomputers the program will run on, main memory requirements, data storage, the method used to store the data, the price of the package and its availability in Australia, for thirty one statistical computer packages.

Appendix B shows that most of the software originated in the United States (67 percent) followed by the United Kingdom (26 percent) and Australia (7 percent). Of all these packages five are distributed in Australia. These are B-STAT +, MASS, MICROSTAT, STATPAC, STATPAK and TRS-80. Two of these MASS and B-STAT + are Australian products and hence supportable in Australia by their developers.

The program language used by at least 70 percent of the programs was Basic with Fortran and Pascal taking 10 percent. Encouragingly, of all the packages looked at, 40% will run of CP/M and may be used on a number of different micro computers. Fifty percent would run on Apple DOS, 10 percent on a TRS-80 and 10 percent on a PEI.

Memory requirements were not available for all the packages. Four packages supplying this information required 48K RAM. One package STATPAC required 128K RAM.

Two aspects of general interest are the type of data that can be used and the method used to store this. This information was not available for all packages but for those it was available all could handle floating point data while only five (50 percent) could handle category data. In regard to storage three packages stored the data in binary form and five used ASCII. Conversion programs may therefore be required if data are to be transferred from one system to another.

The price of the packages tended to vary with the procedures incorporated. The range was from A$32 to A$849 with the majority of packages priced between A$100 to A$300.

MICROSTAT: A REVIEW

Introduction

The previous section provided an overview of some of the statistical packages available for Micro-computers. MICROSTAT is one of these packages. It provides a reasonable array of statistical procedures and is generally available in Australia. This section presents a review of MICROSTAT with the aim of providing an indication of the role of micro-computer statistical packages in transport planning.

The review starts by discussing the procedures available in the package. This discussion will be illustrated by an analysis of data on the speeds of vehicles in residential streets. The data used were extracted from a study by Roycroft and Nicholson (1979). After this discussion a number of considerations that should be taken into account when purchasing a statistical package will be presented. These include the accuracy of the results, the maximum data set, computer process time and the documentation.

Procedures

MICROSTAT (ECOSOFT, 1980) was specifically designed for 8-bit micro-computers operating under CP/M. Its program and data files are held on floppy disks. The programs are menu driven and completely interactive. Output is displayed on a video screen or may be optionally sent to a printer.

Upon entry into the MICROSTAT system the user is provided with a list of fourteen procedures (Fig.1). A number of these procedures are discussed in the following paragraphs. The package was run on a Kaypro II.

The first of the fourteen procedures is DATA MANAGEMENT. This includes thirteen data manipulating routines (Fig. 2), which enable the user to ENTER, LIST, EDIT, DELETE and TRANSFORM data. The user can also COPY, SORT, DESTROY and RENAME files. Details of the data sets and programs on each directory can be obtained.

-------------------------------- MICROSTAT --------------------------------

A. DATA MANAGEMENT SUBSYSTEM	H. REGRESSION ANALYSIS
B. DESCRIPTIVE STATISTICS	I. TIME SERIES ANALYSIS
C. FREQUENCY DISTRIBUTIONS	J. NONPARAMETRIC STATISTICS
D. HYPOTHESIS TEST; MEAN	K. CROSSTAB / CHI-SQUARE TESTS
E. ANALYSIS OF VARIANCE	L. PERMUTATIONS / COMBINATIONS
F. SCATTERPLOT	M. PROBABILITY DISTRIBUTIONS
G. CORRELATION MATRIX	N. HYPOTHESIS TEST; PROPORTIONS
	O. [TERMINATE]

Figure 1 Main Menu for MICROSTAT

-------------------------- DATA MANAGEMENT SUBSYSTEM --------------------------

A. ENTER DATA	H. DELETE CASES
B. LIST DATA [DEFAULT]	I. VERTICAL AUGMENT
C. EDIT DATA	J. SORT
D. RENAME FILE / EDIT HEADER	K. RANK-ORDER
E. FILE DIRECTORY	L. LAG TRANSFORMATIONS
F. DESTROY FILES	M. COPY FILES
G. MOVE / MERGE / TRANSFORM	N. [TERMINATE]

Figure 2 Data Management System Menu

The remaining thirteen procedures outlined in Fig. 1 are the statistical procedures available in MICROSTAT. The first step in the present analysis of speeds in residential streets is to use the DESCRIPTIVE STATISTICS. Figure 3 presents a typical output. A mean speed of 51.2 km/hr with a standard deviation of 6.3 km/hr was found. The upper end of the range of speed shows that some people are exceeding the speed limit. Information on the coefficient of variation, skewness and kurtosis could be obtained if required.

--------------------- DESCRIPTIVE STATISTICS ---------------------
HEADER DATA FOR: B:SPEED1 LABEL:
NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

SPEED STUDY

NO.	NAME	N	MEAN	STD. DEV.	MINIMUM	MAXIMUM
1	SPD	74	51.2243	6.3284	31.6000	64.8000
2	VEH	74	5.4054	2.3982	1.0000	12.0000
3	OVEH	74	2.8108	1.9909	.0000	9.0000
4	FLO	74	1.0811	.4796	.2000	2.4000
5	OFLO	74	.5622	.3982	.0000	1.8000
6	LOC	74	14.2014	18.1360	.0000	75.0000
7	INT	74	9.1424	5.0036	2.3900	12.8000
8	WDH	74	5.2965	1.0377	2.8000	6.4500

Figure 3 Descriptive Statistics for speed in residential streets study.

(Variable names : SPD = Speed, VEH = Number of vehicles travelling in direction considered, OVEH = Number of vehicles travelling in opposite direction, FLO = Flow in direction considered, OFLO = Flow in opposite direction, LOC = percentage local traffic, INT = Number of intersections in street, WDH = width of road).

To look at the general distribution of speeds the FREQUENCY procedure can be used. Figure 4 shows that five vehicles were travelling at a speed greater than 60 km/hr and the distribution of speed appeared to be approximately normal.

```
=====CLASS LIMITS====  FREQUENCY ..............................
    30.00 <    32.00          1 |==
    32.00 <    34.00          0 |
    34.00 <    36.00          1 |==
    36.00 <    38.00          1 |==
    38.00 <    40.00          1 |==
    40.00 <    42.00          2 |====
    42.00 <    44.00          1 |==
    44.00 <    46.00          5 |============
    46.00 <    48.00          6 |==============
    48.00 <    50.00          9 |=====================
    50.00 <    52.00         12 |=============================
    52.00 <    54.00         10 |========================
    54.00 <    56.00          8 |===================
    56.00 <    58.00          6 |==============
    58.00 <    60.00          6 |==============
    60.00 <    62.00          4 |=========
    62.00 <    64.00          0 |
    64.00 <    66.00          1 |==
```

Figure 4 Distribution of vehicle speeds

The data on the speeds of vehicles was collected from three different streets with different road widths and number of intersections. The HYPOTHESIS TESTING and ANALYSIS OF VARIANCE procedures enable the analyst to compare the speeds on the different streets Figure 5 shows that the speed on street one is significantly higher than the speed on street 2.

The next stage in an analysis is to look at the relationship between each of the variables. SCATTERPLOT allows a pictorial view of the data to be obtained. Figure 6 shows such a plot for vehicle speed and flow rate of vehicles. The only inconvenience found in this plot was the incomplete axis labelling. CORRELATION MATRIX then allows the user to obtain an indication of the relationship between each attribute. Figure 7 presents these correlations. A covariance matrix can also be determined.

The analysis procedure outlined above was conducted in less than half an hour elapsed time. Analyses of this type using mainframe computers are likely to be considerably more time consuming.

------------------- HYPOTHESIS TESTS FOR MEANS ------------------
HEADER DATA FOR: B:SPEED1 LABEL:
NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

DIFFERENCE BETWEEN TWO GROUP MEANS: SMALL SAMPLE

SPEED STUDY

	GROUP 1	GROUP 2
MEAN =	56.1077	46.7792
STD. DEV. =	3.8290	6.3287
N =	26	24
CASES =	1 TO 26	27 TO 50

DIFFERENCE = 9.3285
STD. ERROR OF DIFFERENCE = 1.4662

T = 6.3624 (D.F. = 48) VARIABLE: SPD

PROB. = .0001

--------------------- ANALYSIS OF VARIANCE ---------------------
HEADER DATA FOR: B:SPEED1 LABEL:
NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

ONE-WAY ANOVA

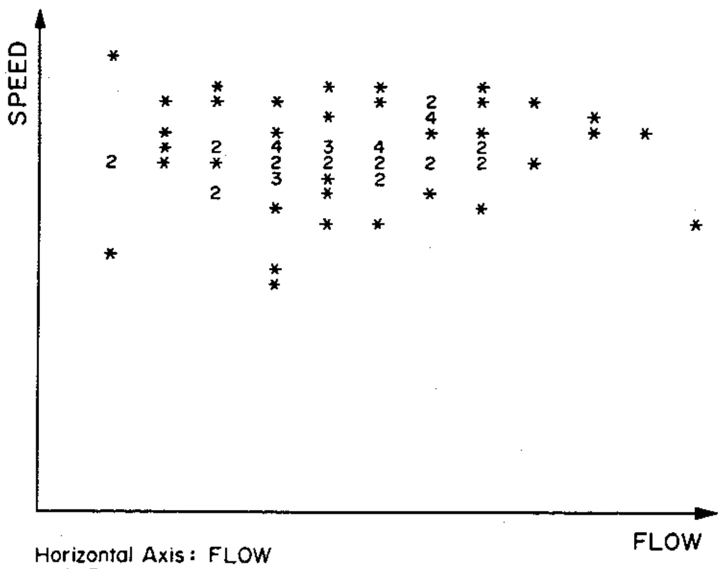
SPEED STUDY

GROUP	MEAN	N
1	56.108	26
2	46.779	24
GRAND MEAN	51.630	50

VARIABLE 1: SPD

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RATIO	PROB.
BETWEEN	1086.027	1	1086.027	40.481	.0001
WITHIN	1287.758	48	26.828		
TOTAL	2373.785	49			

Fig. 5 Comparison of mean speeds at each site



Horizontal Axis: FLOW
Left Endpoint: 0 Right Endpoint: 2.5

Vertical Axis: SPEED
Lower Endpoint: 0 Upper Endpoint: 70

Header Data for: B:SPEED LABEL:
Number of Cases: 74 Number of Variables: 8

Figure 6 SCATTERPLOT of vehicle speed and flow rate

CORRELATION MATRIX

HEADER DATA FOR: B:SPEED1 LABEL:-
NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

SPEED STUDY

	SPD	VEH	OVEH	FLO	OFLO	LOC	INT	WDH
SPD	1.000							
VEH	.125	1.000						
OVEH	.235	.292	1.000					
FLO	.125	1.000	.292	1.000				
OFLO	.235	.292	1.000	.292	1.000			
LOC	-.217	.185	-.018	.185	-.018	1.000		
INT	-.572	-.398	-.428	-.398	-.428	-.262	1.000	
WDH	.181	-.226	.070	-.226	.070	-.063	-.079	1.000

Figure 7 CORRELATION between variables

An overall relationship between the speed and the attributes of the traffic and street was obtained using REGRESSION ANALYSIS. Figure 8 presents the results. It can be seen that although there is a resonable relationship, several variables had insignificant parameters at the 5% level. These could be removed and another regression equation obtained. The REGRESSION ANALYSIS also allows a residual plot to be prepared. This plot may provide an indication of appropriate data transformations.

---------------------- REGRESSION ANALYSIS ----------------------

HEADER DAIA FOR: B:SPEED1 LABEL:

NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

SPEED SIUDY

INDEX	NAME	MEAN	SID.DEV.
1	VEH	5.405	2.398
2	OVEH	2.811	1.991
3	FLO	1.081	.480
4	OFLO	.562	.398
5	LOC	14.201	18.136
6	INT	9.142	5.004
7	WDH	5.296	1.038
DEP. VAR.:	SPD	51.224	6.328

DEPENDENI VARIABLE: SPD

VAR.	REGRESSION COEFFICIENI	SID. ERROR	I(DF= 68)	PROB.	PARIIAL r^2
FLO	-.5837	1.3237	-.441	.6606	.0029
OFLO	-1.1478	1.5680	-.732	.4667	.0078
LOC	-.1366	.0320	-4.272	.0001	.2116
INT	-.9049	.1339	-6.759	.0001	.4018
WDH	.5774	.5557	1.039	.3025	.0156
CONSTANI	59.6555				

SID. ERROR OF ESI. = 4.689
R SQUARED = .489
MULTIPLE R = .699

ANALYSIS OF VARIANCE IABLE

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RAIIO	PROB.
REGRESSION	1428.248	5	285.650	12.990	.0001
RESIDUAL	1495.348	68	21.990		
IOTAL	2923.596	73			

Figure 8 REGRESSION ANALYSIS of variables influencing speed in residential streets

The last two studies that were carried out on the data were a CROSSTAB/CHI-SQUARED to test if the flow in the west and easterly direction were related (Appendix C) and a TIME SERIES STUDY of the changes in the flow rate with time (Appendix D).

This total analysis took a little over half an hour, once the data had been input. The interactive nature of the packaged provided an ability to interrogate the data in a systematic and informative manner. This aspect of the micro-computer appeared to be a very useful aid to the analyst.

Accuracy

The accuracy of statistical procedures is an important overall consideration, for if the package does not provide consistently correct answers it is likely to fall into disrepute and disuse. In this study the analysis described in the previous section was repeated using SPSS. A comparison between the SPSS and MICROSTAT showed no appreciable differences in the results, providing some evidence that MICROSTAT provides acceptable results. Further evidence will only come with wider applications of the package.

Data storage

The maximum data set that MICROSTAT can analyse is a function of the computer system used to run the program and its disk capacity. In this study a KAYPRO II was used to run the program. The KAYPRO II has 56 Kilobytes of useable random access memory (RAM). A further 8 kilobytes are used in the control programs (CP/M). MICROSTAT requires 48 kilobytes, therefore there are 8 kilobytes available for storage of extra data in RAM. The statistical procedure that appears to use most memory is REGRESSION. In the KAYPRO II the useable RAM for this procedure can hold approximately 1000 data elements. If the data exceeds this amount MICROSTAT reads the data into memory from the disk. The limit on the total size of the data set that can be analysed is therefore determined by the disk capacity.

The KAYPRO II has two disk-drives which read two 5.25 inch single sided double density disks. Each disk can hold 195 kilobytes. MICROSTAT requires 115 kilobytes of storage and can therefore be stored on one disk leaving 80 kilobytes for data. It is more common in micro-computers with two disks to store the data on the second disk, hence providing a capacity of 195 kilobytes for the data. Approximately 20,200 pieces of data can be stored on the second disk. With a maximum of 30 variables then it is possible to have a maximum of approximately 670 cases for each variable when using KAYPRO II.

Computer Time

Possibly the major advantage associated with a micro-computer is its ease of access. Each individual can purchase a system like the KAYPRO II for about $2,500 and it can be used wherever there is a power point. Individual ownership removes the need to wait in queues behind other computer users, the typical situation in a mainframe system. Micro's are therefore well suited to an interactive format where the computer uses a series of "menus" to ascertain the users requirements. This inturn provides the user with a direct feel for what the data are saying. The limit on the extent of such exploratory studies is likely to be influenced by the time taken to run the statistical procedures.

An indication of the time taken to run each of the procedures was obtained by analysing a series of test data sets by MICROSTAT on the KAYPRO II. To present these findings the general flow of MICROSTAT will be followed.

The first procedure in MICROSTAT is a DATA MANAGEMENT SYSTEM. This allows a number of data manipulations to be carried out (Fig. 2). The Data ENTRY and EDIT routines are interactive and their time of operation depends on the speed of the operator. The LIST data, file RENAME, file DIRECTORY and DESTROY files routines are fast needing only a few seconds to run. The MERGE/MOVE/TRANSFORM and DELETE routines can however become quite slow. Both these routines require the micro to consider each case in turn. Transformations require somewhere in the order of two minutes per hundred cases. Deleting cases requires 2 minutes per hundred cases if it is not necessary to change the case number. If it is necessary to change the case number this time can rise to 3.5 minutes per 100 cases. This is slow and it is unlikely that the user would sit by the machine for the time to consider data sets of over 200-300 cases.

The COPY file command should also be discussed. This procedure copies the contents of one file to another. It does this by inputing the entire file into core then transferring it to the new location. Hence it can handly only approximately 1000 pieces of data.

An alternative to MOVE and COPY is to use the system utilities (e.g. those in CP/M) for transferring data files. This may offer considerable advantages in those situations where it applies.

The remaining thirteen procedures (Fig. 1) in MICROSTAT are statistical procedures. Many of these procedures consider only one or two variables (FREQUENCY distribution, HYPOTHESIS TESTING, SCATTERPLOTS). These very useful procedures take only a minute or so to consider several hundred data points. To provide an indication of the time involved in running the slower routines consider the three commands

DESCRIPTIVE STATISTICS
CORRELATION MATRIX
REGRESSION ANALYSIS.

As discussed earlier the DESCRIPTIVE STATISTICS provide measures of the mean, standard deviation etc. This routine was found to take 0.16 second per piece of data. Hence a data set that contains 500 cases and 20 variables will require 26 minutes to run. The times for the CORRELATION ANALYSIS to run are illustrated in Fig. 9. It can be seen that for 16 variables and 500 cases the program would require 26 minutes to run. The last routine to be considered is REGRESSION analysis. Figure 10 shows that for Multiple Linear Regression the time taken to run the program increases linearly with increases in cases and quadratically with increases in the number of parameters. A data set of 500 cases and 8 variables would take 26 minutes to run.

It is apparent from this discussion that several of the routines in MICROSTAT take a reasonable length of time to analyse medium size data sets in the operating mode used. Data sets of the order of 2-300 cases and 5 to 10 variables appear to provide an upper limit to the effective use of the interactive capabilities of the system. However if the user is content to leave the system and let it run then it is possible for MICROSTAT to process reasonably large data sets. Further, other users using different microcomputers have found large differences in MICROSTAT execution times for similar jobs. The data presented for the KAYPRO II should only be seen as a guide.

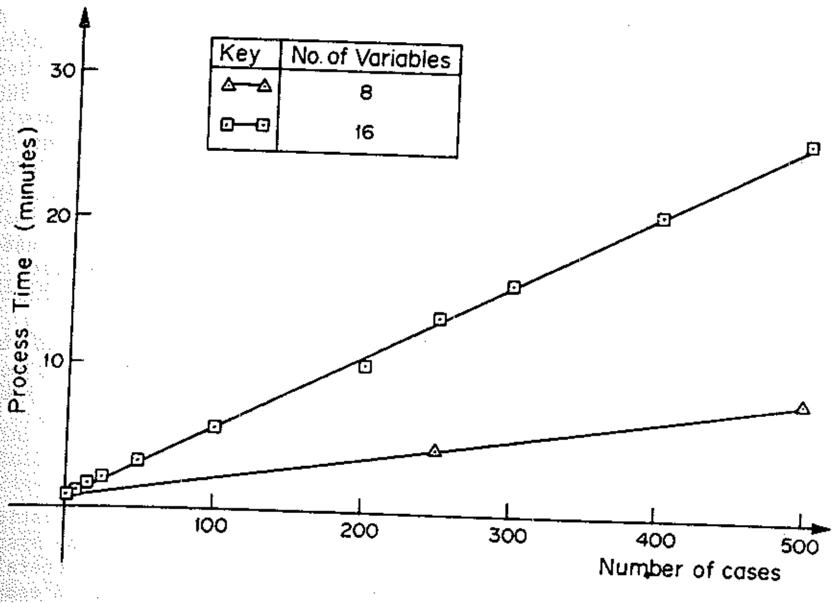


Figure 9. Computer Running Time for Correlation Analysis

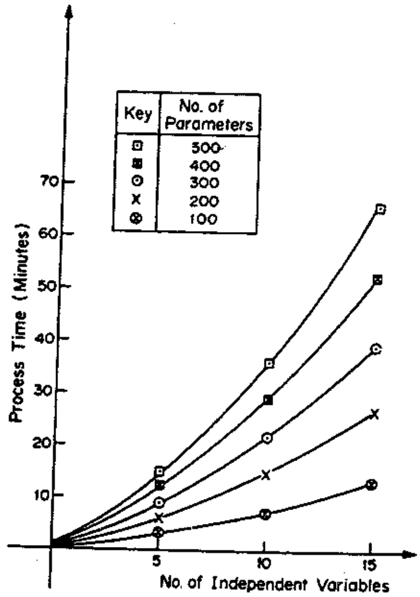


Figure 10. Computer Running Time for Regression Analysis

Documentation

The final aspect that may influence the purchase and usage of MICROSTAT is the documentation associated with the program. Although MICROSTAT operates in an interactive mode and the questions asked are direct and understandable, it is necessary to expand on these questions to gain insight into the total capabilities of the program. MICROSTAT documentation consists of about 45 pages of description on the usage of the programs and a further 53 pages of appendices which describe a number of example runs (ECOSOFT, 1980). The documentation is therefore aimed primarily at running the program and the authors found it concise and clear. The package includes the test data sets analysed in the appendices for checking purposes by the individual user. Analysing these test data sets helps the user to learn the capabilities of MICROSTAT.

For the user without a reasonable background in statistics the documentation is not adequate. It is however reasonable to expect that most users would seek professional help in any attempted analysis beyond their statistical knowledge.

Remarks

MICROSTAT was found to present a considerable array of statistical procedures of use to the transport planner. The main limitation of the system appeared to be the execution time involved in particular tasks. Generally sample sizes of 200-300 appeared to enable full use of the interactive component of the program. However once the sample sizes rose above this the machine took a considerable time to carry out tasks.

MICROSTAT is therefore likely to be best used as a tool for small to medium size data sets, leaving the mainframe to carry out the full computational load. In such a situation MICROSTAT offers a convenient means of data entry and editing, although the transfer of data require translation programs to convert data from binary decimal to ASCII and vice versa. The reason for this is that MICROSTAT uses its own binary data files. Such translation programs are not supplied with MICROSTAT and the user must provide his own.

CONCLUSIONS

Micro-computers have grown rapidly in power and availability over the last five years. Their potential as a tool for transport planners is increasing daily and the full extent of their possible usefulness is unknown but obviously considerable. This paper reviewed some statistical packages for currently available systems.

There are many packages available. These packages have most statistical procedures needed for general computer analysis but lack some of the more specific tools like logit analysis, probit analysis and factor analysis commonly used in transport planning. Such procedures are being incorporated into a number of packages.

A detailed study of one of the more well used packages (MICROSTAT) revealed that the package could analyse data sets of about 5000 entries in a reasonable time but became time consuming for larger data sets.

The interactive capabilities of the micro-computer make MICROSTAT very useful for interrogation of small to medium size data sets. Larger data sets may need transfer to mainframes after preliminary analysis on micro-computers, so that the broader range of statistical procedures available on such systems can be used. Further developments in microcomputers may soon result in a change in this conclusion!

ACKNOWLEDGMENTS

The authors wish to thank Alan Miller, CSIRO, Division of Mathematics and Statistics for his discussion on many of the packages referred to in this paper. The views expressed in the paper are however those of the authors.

REFERENCES

Abelson, P.W. and Baker, S. (1982). "Household car ownership and mode choice for the journey to work," Forum Paper, 7th Australian Transport Research Forum, Vol 2, pp 607-623.

Aplin, W.N. (1983). "Analysis of repair and maintenance costs for log vehicles", Forum Paper, 8th Australian Transport Research Forum, Vol 3, pp 244-264.

Bradley, R. and Holsman, A. (1983). "Some observations on the impact of main road traffic noise in Sydney", Forum Paper, 8th Australian Transport Research Forum, Vol. 3, pp 98-118.

Brown, H.P. (1982). "Aggregate attraction measures for disaggregate destination choice models", Forum Paper, 7th Australian Transport Research Forum, Vol. 1, pp 137-150.

Carterette, E.C. and Friedman, M.P. (Eds.) (1974). "Handbook on perception Vol. 11: Psychophysical judgment and measurement", (Academic Press: New York)

Daganzo, C.F. (1980). "Calibration and prediction with random utility models: Some recent advances in unresolved questions", Resource Paper for 4th International Conference on Behavioural Travel Modelling, Eibsee, West Germany.

Dumble, P.L. (1979). "Urban freight and land use interaction", Forum Paper, 5th Australian Transport Research Forum, pp 422-440.

ECOSOFT (1980). "MICROSTAT: User Manual."

Faulkner, H.W. and Nelson, R.M. (1983). "Job accessibility and unemployment in metropolitan Melbourne: A spatial analysis", Forum Paper, 8th Australian Transport Research Forum, Vol. 3, pp 70-97.

GENSTAT (1975). "The GENSTAT language: GENSTAT User Guide No. 1", Statistics Department, Rothamsted Experimental Station, Harpenden, Herts.

Hensher, D.A. and Manefield, T. (1982). "A structural logit model of automobile acquisition and type choice", Forum Paper, 7th Australian Transport Research Forum, Vol 2, pp 625-654.

IMSL (1982). "International mathematical and statistical library, INC: Reference Manual", (IMSL; Texas)

Neffendorf, N. (1983). "Statistical packages for microcomputers: A listing", The American Statistician, Vol. 37, No. 1, pp 83-86.

Nicolaidis, G.C. (1975). "Quantification of the comfort variable", Transportation Research, Vol 9, p 55.

Roycroft, D.C. and Nicholson, A.M. (1979). "Factors affecting vehicle speeds in residential streets", Monash Civil Engineering Working Paper 79/10.

Richardson, A.J., Kok, Y.S. and Vandebona, U. (1983). "Microcomputers and computer graphics in transport planning and design", Forum Paper, 8th Australian Transport Research Forum, Vol 1, pp 323-343.

SPSS (1972). "Statistical package for the social sciences: Users Manual", (McGraw-Hill; New York)

Stopher, P.R. and Meyburg, A.H. (Eds.) (1976). "Behavioural Travel Demand Models", (Lexington Books: Lexington).

Taylor, M.A.P. (1978). "Traffic distribution in a residential cell; A case study", Forum Paper, 4th Australian Transport Research Forum, pp 567-590.

Young, W. (1979). "Modelling residential location choice: A review of some results", Forum Paper, 5th Australian Transport Research Forum, pp 373-393.

Young, W. and Ogden, K.W. (1983). "Analysis of freight-facility location choice using an Elimination-by-Aspects Model", Forum Paper, 8th Australian Transport Research Forum, Vol. 3, pp 265-281.

Young, W. and Richardson, A.J. (1978). "Residential preference: A perceptual view of Melbourne", Forum Paper, 4th Australian Transport Research Forum, pp 591-622.

Appendix A STATISTICAL PROCEDURES IN MICRO-COMPUTER STATISTICAL PACKAGES

	ABSTAT	ADVANCED MATHEMATICS	AIDA	AIDA-I	A.STAT	B.STAT +	DAISY	DYNACOMP	INTERSTAT	MASS	MATHSTAT	M D A	MICROSTAT	P D A	SAM	SL-MICRO	SPP	STATISTICAL MICROPROGRAMS	STATISTICS PAC	STATISTICS PACKAGE (LOM)	STATISTICS PACKAGE (OLD)	STATISTICS-3	STATPAC	STATPAK	STEP WISE REGRESSION	TECPACS	TRS-80 STATISTICS
Data Investigation																											
Descriptive Statistics	x		x		x	x	x		x	x	x	x	x		x	x	x	x	x	x	x	x	x	x		x	x
Frequency Distributions	x		x	x	x	x			x	x	x	x	x		x	x	x				x	x	x	x		x	x
Scatterplots	x	x	x	x		x	x		x	x		x	x	x	x		x						x	x			
Cross Tabulations	x				x	x				x	x	x	x		x	x	x					x	x	x			
Hypothesis Testing																											
Chi Squared	x					x			x	x		x	x		x		x			x		x	x	x		x	x
Non-Parametric							x	x	x	x			x				x						x	x			
Difference in Means	x		x				x		x	x	x		x		x		x	x			x	x	x	x			x
Correlation	x	x	x	x	x	x	x		x	x	x	x	x		x	x	x	x		x	x	x	x	x	x	x	x
Analysis of Variance	x		x			o	x	x	x	x	x	o	x			o	x	x		x	x		x	x			x
Distribution Fitting						x																					
Relationship Determination																											
Multiple Linear Regression	x	x	x	x	x	o	x	x	x	x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	x
Stepwise Regression											o		x											x			
Discriminant Analysis										x					x												
Logit Analysis											o	x															
Probit Analysis											o																
Log Linear Analysis										x																	
Time Series Analysis				x			x			x			x														x
Data Reduction																											
Principal Component Analysis								x			x																
Factor Analysis								x			o	o			x												
Canonical Correlation											o																
Multidimensional Scaling											o																

x Available procedures
o Procedures to be incorporated by early 1984

APPENDIX B GENERAL INFORMATION ON STATISTICAL PACKAGES.
[This Table is an extension of a table presented by Neffendorf (1983)]

Package Name	Supplier	Language	Micro-Computers	Memory Req'd	Data Form	Data Storage	Price	Remarks
Abstat	Anderson-Bell 2916 So.Stuart St. Denver, Colorado, 80236	NK	CP/M	NK	NK	NK	US$40	*
Advanced Mathematics	Great Northern Comp. Services 116 Low Lane Horsforth, Leeds LS18 5PX	Basic	Apple	NK	NK	NK	£150	*
AIDA	Action-Research Northwest 1142 Marine View Drive, S.W., Seattle, Washington 98146	Basic	Apple	NK	NK	NK	US$235	*
AIDA-I	Appli-Tech Software Services Broad Oak, Accrington BB5 2DJ	Pascal	Apple IBM-PC ACT-Sirius	NK	Floating Point	Binary	£450	Δ
A.STAT	Rosen Grandon Assoc. 296 Peter Green Rd Tolland Connecticut 06084	Basic	CP/M Apple PET	NK	NK	NK	US$124	*
B-STAT +	Micro Stat Software P.O.Box 125 Queanbeyan 2602	C & Fortran	CP/M	NK	Floating Point	ASCII	$175	Dev. in Aust.
Conduit	Conduit 100 Lindquist Ctr. Univ. of Iowa, P.O.Box 388 Iowa City, IA 52244	Basic	Apple	NK	NK	NK	US$100 approx.	*
DAISY	Rainbow Comp. Inc. 19517 Bus. Centre Drive, Northridge California 91324	NK	Apple	48K	Floating Point	NK	US$80	Δ

Package Name	Supplier	Language	Micro-Computers	Memory Req'd	Data Form	Data Storage	Price	Remarks
Dynacomp	Dynacomp Inc. 1427 Monroe Ave Rochester New York 14618	Basic	Apple Atari TRS-80 PET	NK	NK	NK	Reg $119 $119 Δ ANOVA US$119	
INTER-STAT	Great Northern Comp. Services 116 Low Lane Horsforth, Leeds LS 18 5PX	Basic	Apple	48K	NK	NK	£150	Δ
MASS	Westat Assocs. 60 Bruce St. Nedlands,WA 6009	Pascal	CP/M	NK	NK	NK	NK	Δ
MDA	Cambridge Inf. International 238 Main Street, Suite 310, Cambridge, MA02142	Basic	CP/M	NK	Floating Point & Charac.	NK	US$349	Δ
Math Stat	Mathematica Policy Research Inc P.O. Box 2393 Princeton, NJ 08540	Basic	IBM/PC	NK	Floating Point & Alpha- numeric	Binary	US$750	Δ
Microstat	Ecosoft P.O.Box 68602 Indianapolis Ind. 46268	Basic	CP/M	48K	Floating Point	Binary	A$350	Dist. in Aust.
Personal Data Analysis	Personal Comp Ltd. 220-6 Bishopsgate London EC2	Basic	Apple	NK	NK	NK	7 mods £75-125 each	*
SAM	International Software (UK) PO Box 160 Welwyn Gdn. City Herts AL8 6TQ	Basic	Apple PET CP/M	NK	Floating Point & Category	ASCII	£335	Δ
SL-MICRO	Questionnaire Service Company Box 23056 MI 48909	Basic	CP/M	48K	Floating Point & Category	ASCII	US$280	Δ

Package Name	Supplier	Lang-uage	Micro-Computers	Memory Req'd	Data Form	Data Storage	Price	Remarks
SSP	Patrick Royston 85 Canfield Gdns London, NW6 3EA	Basic	PET	NK	NK	NK	£350	*
Statistical Analysis	Aerocomp Inc. 2203 Harbor Bvd Costa Mesa California 92626	NK	CP/M	NK	NK	NK	NK	*
Statistical Analysis System	H & H Associates Box 2663 Renton Washington 98055	NK	CP/M	NK	NK	NK	NK	*
Statistical Micro Programs	Statistical MicroPrograms US address unknown	Basic	Apple	NK	NK	NK	US$150	
Statistics	Basic Business Software Box 2032 Salt Lake City, UTAH 84110	NK	CP/M	NK	NK	NK	NK	*
Statistics	Creative Discount Software 256 S Robertson Suite 2156 Beverly Hills California 90211	Basic	TRS-80 Apple	NK	NK	NK	US$100	Δ
Statistics Package	Lombardy Computers 121 High Street Berkhampstead Herts HP4 2DJ	Basic	Apple	NK	NK	NK	£122	*
Statistics Package	Old Bird Software John C. Dvorak 704 Solano Ave Albany, California 94706	Assem-bler	CP/M	NK	NK	NK	US$50	*
Statistics -3	Edu-Ware Services Inc. 22222 Sherman Way Canoga Park CA 91303	Basic	Apple	48K	NK	NK	US$30	Δ
STATPAC	Wallonick Assoc. 5624 Girard Av.Sth Mineapolis, 55419	Assem-bler	IBM/PC	128K	Floating	NK	A$400	Dist.

Supplier	Language	Micro-Computers	Memory Req'd	Data Form	Data Storage	Price	Remarks
Northwest Analytical Inc. P.O.Box 14430 Portland Oregon 97214	Basic	CP/M	48K	NK	ASCII	US$500	Dist. in Aust.
Apple dealers	Basic	Apple	NK	NK	NK	£77	*
BHRA Fluid Engineering Bedford MK43 OAJ	Basic	PET Apple	NK	NK	NK	Sev. modules £80-100	*
TRS-80 Applics. Software Sourcebook Tandy Radio Shack Stores£3.95	Basic	TRS-80	NK	NK	NK	A$60	Dist. in Aust.

t known
d not answer letter of communication
er support by letter or telephone.

APPENDIX C CHI-SQUARED TEST ON DIRECTIONAL FLOWS FOR ONE STREET

------------------ CROSSTAB / CHI-SQUARE TESTS ------------------
SPEED STUDY

OBSERVED FREQUENCIES

	1	2	TOTAL
1	9	9	18
2	8	4	12
3	7	6	13
4	7	4	11
5	10	6	16
6	8	2	10
7	6	1	7
8	7	6	13
9	6	4	10
10	5	6	11
11	9	2	11
12	3	6	9
13	6	0	6
14	6	2	8
TOTAL	97	58	155

CHI-SQUARE = 15.597, D.F. = 13, PROB. = .2716

APPENDIX D TWENTY MINUTE MOVING AVERAGE OF FLOW RATE

--------------------- TIME SERIES ANALYSIS ---------------------

SPEED STUDY

HEADER DATA FOR: B:SPEED1 LABEL:
NUMBER OF CASES: 74 NUMBER OF VARIABLES: 8

	FLO	4 TERM MOVING AVG.
1	1.80	
2	1.60	
3	1.40	
4	1.40	1.55
5	2.00	1.60
6	1.60	1.60
7	1.20	1.55
8	1.40	1.55
9	1.20	1.35
10	1.00	1.20
11	1.80	1.35
12	.60	1.15
13	1.20	1.15
14	1.20	1.20