Statistical programs for Commodore computers

DAVID J. PITTENGER and MILTON H. HODGE
University of Georgia, Athens, Georgia

In research and education, microcomputers are frequently used for statistical analysis, but many existing programs have disparate data entry, storage, and analysis routines. The analyst must reenter the same information for each analysis, often in a different format with a different entry routine, which may confuse even experienced users. The usual solutions to these problems are to organize a set of programs around an easy-to-use menu and to develop an entry-and-storage procedure that creates a data file compatible with all the statistical procedures. In this paper, we describe a set of programs which avoids these problems and offers an efficient approach to a variety of statistical analyses.

The package of programs has several significant features: (1) The programs perform many common statistical analyses; however, they do assume that the user is familiar with statistical concepts. (2) The programs are easy to use because they require little knowledge of programming or operation of computers and, consequently, are well suited for student use. (3) All of the longer statistical programs have been converted to machine code to make them run quickly and to reduce computation time with large data sets. (4) Two utility programs, Data Manager and Disk Operations, facilitate and ease use of the statistical analyses. (5) Although designed expressly for Commodore computers, the programs are written entirely in BASIC, making them easily adaptable to other machines which run BASIC programs. (6) Working memory (RAM) of the Commodore computers is not as large as RAM of newer machines, but it is sufficient for the analysis requirements of many investigators and many student projects. (7) Because a Commodore 64 system (C-64 computer, 1541 disk drive, 1702 color monitor) can be purchased for less than \$500 (1985), the programs offer an inexpensive source of considerable computing power.

Overview of Package

Hardware Requirements. Written in Commodore BASIC, all of the programs run on Commodore PET/CBM 4032 and 8032 computers and the Commodore 64. In addition to the computer and an appropriate monitor, a disk drive and a printer are required. The programs are stored on a diskette, formatted to operate with

4040, 2031, and 1541 disk drives and other non-Commodore drives compatible with the Commodore system. Unless we state otherwise, the results of all computations are listed on a printer.

Accuracy. All computational devices are subject to some degree of inaccuracy. In general, an error occurs when the magnitude of a number exceeds numerical capacity of the computer. Computers typically convert these large numbers into scientific notation, but many significant digits of the original numbers may be lost via truncation. The loss of digits is especially likely when numbers are squared and then summed, as in calculations of variance, standard deviation, and covariance. To circumvent this problem, all of our programs use updating routines (see Lachenbruch, 1983), which minimize the loss of significant digits, and thereby preserve the accuracy of the original scores.

Programs

Master Menu. As noted, the programs are organized around a Master Menu that displays all programs in the package. Typing the number of a program loads the program, ready to use, into the computer. The Master Menu is automatically reloaded at the completion of each program. The Master Menu also contains routines to change the screen and letter colors of the monitor for C-64 systems.

Data Manager. The Data Manager program is used to create data files, which then serve as the input for most of the statistical analyses in the package. To create a data file, the user specifies the number of treatment groups (cells) in the experimental design and specifies whether the number of observations is equal or unequal across groups. After the data have been entered, they can be proofread and revised as needed (errors may be corrected and unwanted scores deleted). The data then can be printed and saved on a diskette. The data file can be recalled at a later time for further proofing and revision. Possible revisions include adding/deleting data to/from existing groups or adding new groups to the file. The revised file can replace the old file, or it can be saved as a new file on the diskette.

Disk Operations. The Disk Operations program performs a variety of typical disk housekeeping operations, such as formatting and validating data disks used with the programs, reviewing the catalog of files on a data disk (with an option for printing the directory on the printer), scratching or erasing unwanted files, and copying files from one disk to another disk. Numerous warning messages help prevent the accidental erasure of data files. This program, along with Data Manager, facilitates the creation and use of the data files employed in the analysis programs.

Descriptive Statistics. This program calculates a variety of descriptive statistics for each group included in a

Development of the programs was supported in part by the University of Georgia Office of Instructional Development. Reprint requests should be addressed to either author, Department of Psychology, University of Georgia, Athens, GA 30602.

data file. A menu displayed on the monitor permits the user to select particular measures or request the complete set of statistics. The program calculates: a frequency distribution, including frequency of each score, cumulative frequencies, percent of each score of total scores, and standard deviate (z) of each score; harmonic and geometric means; median; mode; variance and standard deviation; standard error of the mean; range; quartiles and quartile range; coefficient of variation; and skewness and kurtosis indices. The sum of scores, sum of squares, and arithmetic mean are calculated and printed by default for each group.

Correlation/Regression. The Correlation/Regression program computes all possible Pearson correlation coefficients (r) among the groups or treatments contained in a data file and performs bivariate and multiple linear regression analyses for any combination of the groups. The correlation routine calculates and prints out the correlation coefficient, the t -ratio for the correlation, and the sum of cross products for each combination of groups. At the user's option, the mean, standard deviation, sum of scores, and sum of squares for each group can also be printed. The regression routine permits the user to designate any group as the dependent variable and any number of the remaining groups as independent variables. The printout of the analysis includes the estimated intercept of the regression equation, the beta weights of each independent variable, and R^2 . Another printout summarizes an analysis of variance of the regression model. The intercept and beta estimates are derived from the inverse of the correlation matrix R (Pedhazur, 1982). Although Commodore computers do not support double-precision mathematical functions, our tests (using data suggested by Longley, 1967) indicate that the computations are relatively accurate when there is a high degree of multicollinearity among the independent variables. (The largest percentage error was within .001 of Longley's calculation.)

Data Transformer. This program is a collection of short routines that perform various mathematical operations on the contents of a data file. The program can add, subtract, multiply, and divide by a constant; square each score; perform square root, log, and reciprocal transformations; test whether any of the transformations improve the homogeneity of variance of the scores (Kirk, 1982; Winer, 1971); and convert the scores to z scores.

Analysis of Variance. Based in part on a program by Lane (1981), this program performs an analysis of variance on experimental designs based on one to nine factors, with and without repeated measures on each factor. Some representative designs that can be evaluated are completely randomized one-factor designs, randomized block designs, completely randomized factorials, and mixed factorials. In designs with unequal subjects in cell, the program performs an unweighted means analysis (Kirk, 1982; Winer, 1971).

The program prints tables that summarize the highest order interaction, subject totals across repeated trials of

within-subject factors, the main effects of all factors, and the results of the analysis of variance.

Significance Tester. Significance Tester, derived from a similar program by von Collani (1983), calculates the exact probability of z scores, correlation coefficients, t ratios, F ratios, and chi squares. The probabilities are accurate to the fourth decimal place (Ling, 1978). The results of each probability test include p and $1-p$ for both one-tail and two-tail tests, depending on the particular test statistic.

Multiple Comparisons. This program includes two routines that permit evaluation of the specific comparisons associated with a treatment variable based on three or more groups. One routine does an a priori orthogonal comparisons analysis, and the other performs an a posteriori Tukey WSD test. Each test requires input of the group means, the number of observations in each group (equal or unequal), and the appropriate mean square error term from an analysis of variance. For the a priori test, the user must also enter the appropriate orthogonal coefficients and the probability level (.05 to .005) at which the comparisons are to be evaluated. The program computes the critical difference value and tests all comparisons against this value. For the Tukey test, the user must supply the appropriate studentized range statistic (Q) from tables (Kirk, 1982; Winer, 1971).

Random-Number Generator. This program generates up to 6,000 pseudorandom numbers and prints a copy on the printer. The user may specify the number of digits and the range of the digits (e.g., 0-99, 0-999, etc.).

Permutations and Combinations. To calculate the number of permutations and combinations for a set of N events taken K at a time, the user must enter the number of N events, the size of the subset K , and the selection process (with and without replacement of the events in the subset). Five equations, displayed on the monitor, summarize the combinatorial possibilities provided by the program. In addition, the program displays help instructions on the monitor.

Documentation and Availability

A comprehensive user manual has been prepared for the package. Each section reviews the purpose, capabilities, and limitations of each program. Also included is a step-by-step description of the program procedures. A copy of the manual, the program disk, and program listings can be obtained by sending a check or money order (foreign requests) for \$10 US to the authors (payable to the Psychology Department, University of Georgia).

REFERENCES

- KIRK, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.
- LACHENBRUCH, P. A. (1983, November). Statistical programs for microcomputers. *Byte Magazine*, pp. 560-570.
- LANE, D. M. (1981). A general analysis of variance program for microcomputers. *Behavior Research Methods & Instrumentation*, 13, 694.

- LING, R. F. (1978). A study of the accuracy of some approximations for t , χ^2 , and F tail probabilities. *Journal of American Statistical Association*, **73**, 274-283.
- LONGLEY, J. W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *American Statistical Association Journal*, **62**, 819-841.
- PEDHAZUR, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart, & Winston.
- VON COLLANI, G. V. (1983). Computing probabilities for F , T , chi-square, and z in BASIC. *Behavior Research Methods & Instrumentation*, **15**, 543-544.
- WINER, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

(Revision accepted for publication February 2, 1986.)