

COMPUTER TECHNOLOGY

Research Assistant: A statistical package for the Apple II microcomputer

MARLEY W. WATKINS

SouthWest EdPsych Services, Inc., Phoenix, Arizona

and

JOSEPH C. KUSH

Deer Valley Unified School District, Phoenix, Arizona

This article presents a brief description of a statistical package (Research Assistant) designed for Apple II microcomputers. Research Assistant calculates descriptive statistics, chi-square statistics, correlation statistics, a variety of test statistics, one-way analyses of variance, and multiple regression analyses for up to 1,000 subjects with 20 variables per subject. The package is menu driven, and keyboard input is fully interactive, allowing for numerous data management capabilities.

Until recently, researchers who wished to perform statistical analyses were limited to commercial packages available only for mainframe computers. Only within the past several years have statistical packages for microcomputers become available. However, two factors have severely limited their overall success in the marketplace. First, because they are relatively expensive, the cost factor has tended to prohibit the moderate or occasional user from such an investment, and second, the majority of currently available packages have rather extreme hardware and memory requirements that further limit their widespread use.

Recently, Standing (1985) presented a basic statistical package written in Ohio Scientific BASIC, and Pittenger and Hodge (1986) developed a comprehensive but inexpensive statistical package for the Commodore computer. In addition, Butler and Jones (1987) recently reviewed a collection of inexpensive statistical packages for the Apple II microcomputer. Their review described five basic packages and compared these packages' hardware requirements, program capacities, accuracy, editing, and error handling. A wide variety of characteristics and statistical routines were indicated, as was somewhat extreme variability in the cost of the packages. This article describes a similar, inexpensive but comprehensive, statistical package for use with the Apple II microcomputer (see Table 1).

Address correspondence and software requests to Marley Watkins, SouthWest EdPsych Services, Inc., P.O. Box 1870, Phoenix, AZ 85001.

OVERVIEW OF PACKAGE

Hardware Requirements

Research Assistant was written in Applesoft BASIC and requires 48K RAM under the DOS 3.3 operating system. The distributed version has subsequently been compiled for improved operating speed. It operates on Apple II+, IIe, and IIc computers with one or two disk drives. Results are displayed immediately on the monitor and/or may be sent to an optional printer.

Accuracy

The statistical computations carried out by Research Assistant were based upon standard formulas and algorithms (Bruning & Kintz, 1977; Ferguson, 1981; Guilford & Fruchter, 1978; Snedecor & Cochran, 1967). However, the accuracy of any statistical program must be verified before its results are accepted by researchers. Consequently, two sets of data ($N=60$; $N=103$) were submitted to Research Assistant on an Apple IIc, and to the Statistical Analysis System (SAS) (SAS Institute, 1985) and the Statistical Package for the Social Sciences (SPSS) (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) programs operating on an IBM 3081K mainframe computer. Results are identical to three decimal places, and results of selected analyses are contained and presented on the Research Assistant disk. Users should be aware that, with any statistical package, hardware limitations and restricted or extremely large data sets will also affect data accuracy. A comparison of SPSS, SAS, and Research Assistant results obtained from the larger ($N=103$) sample data set is presented in Table 2.

Table 1

Requirements, Features, and Limitations of Research Assistant*

GENERAL FEATURES	
Hardware	
RAM	48K
Minimum Number of Disk Drives	1
80-Column Card	No
Printer	Optional
Cost	\$10
Menus	Yes
Copy Protected	No
List Protected	No
Language	Compiled BASIC
Manual	No
Overall Error Trapping	Yes
STATISTICS	
Descriptive Statistics	
Number of Cases	1,000
Mean	Yes
Median	Yes
Mode	No
Variance	Yes
Standard Deviation	Yes
Minimum, Maximum	Yes
Correlation	
Pearson r	Yes
Spearman rho	No
Number of Cases	1,000
Accuracy	4
Linear Regression	Yes
Number of Cases	1,000
Significance Tests	
t Test	Yes
Independent Groups	Yes
Dependent Groups	Yes
Number of Cases	1,000
Accuracy	4
Mann-Whitney U	No
Wilcoxon	No
ANOVA (One-Way Between)	Yes
Number of Levels	9
Maximum Cases per Level	1,000
Accuracy	4
Chi-Square	Yes
Goodness of Fit	Yes
Contingency	Yes
Number of Levels	9*7
Maximum Frequency	1,000
Accuracy	4
OTHER FEATURES	
Graphics	
Scattergram	Yes
Histogram	No
Data Input	
Disk Files	Yes
Specific to Package	Yes
Compatible Across Routines	Yes
Error Trapping	Yes
Data Editing	
Onscreen	Yes
Cases Numbered	Yes
Transforming Variables	Extensive

PROGRAMS

Instructions

Complete user instructions are contained on the Research Assistant disk. They occupy 26K of memory and can be viewed directly on the monitor or sent to a printer. The program is menu driven. Immediately after initial program loading, the instructions program is accessed, allowing the user the opportunity to read information concerning each specific statistical technique, as well as instructions specific to Research Assistant. Included for each statistical program are introductory explanations of the function, assumptions, and limitations of the statistical technique, as well as text references for more detailed explanations.

Data Management

The Data Management program allows the user to enter data into the computer's memory. The program accommodates up to 1,000 subjects, with up to 20 variables per subject. Data may be input from the keyboard or from disk and may be stored on disk. Data files are Apple random access type. Research Assistant spools all data to the data disk and, therefore, is not RAM bound or susceptible to common causes of program crashes, such as loss due to power failures and so forth. Essentially, its limits are those of the data disk itself. A standard DOS 3.3 disk could therefore contain more than 10,000 data values for a single subject. Hard disk users could analyze hundreds of thousands of data values, depending upon volume size. Keyboard input is fully interactive and allows data editing, data transformation, data to be printed, printing missing data values, and data checking to identify extremely unusual values (i.e., possible erroneous data). Corrections to data on the screen are accomplished via the right and left arrow keys.

Descriptive Statistics

This program calculates means, medians, ranges, standard deviations, variances, standard errors, and measures

Table 2
Accuracy Comparisons of a Large Data Set Across SPSS, SAS,
and Research Assistant Statistical Packages

Statistic	SPSS	SAS	Research Assistant
Mean	95.903	95.9029	95.9029
Median	97.00	97.00	97.00
Variance	156.814	156.814	156.814
Standard Deviation	12.523	12.5225	12.5225
Skewness	-.140	-.140	-.138
Kurtosis	-.499	-.4986	-.532
Standard Error	1.234	1.234	1.234

Note—This data set included 103 subjects. Additional data provided by the Research Assistant package included: name of data file, number of scores, number of missing scores; variable label, sum of scores, sum of squared scores, and high and low scores. Total run time was 32 sec. SPSS = Statistical Package for the Social Sciences. SAS = Statistical Analysis System.

*Characteristics taken from Butler and Jones (1987).

of skewness and kurtosis for each variable. In addition, output also includes the number of valid scores, the number of missing scores, the sum of scores, and the sum of squared scores.

Data Transformations

Data transformations are sometimes required due to non-normally distributed data or violations of certain statistical assumptions. The data transformation program allows the user to transform any variable in the data file. Transformed data is written back to the data file. Data may be transformed to z scores, t scores, or standard scores, or may be transformed to logarithmic, reciprocal, square root, Fisher z , or arc sine distributions.

Correlations

The correlation program calculates Pearson product-moment, partial, and multiple correlation coefficients. Correlations may be calculated for any/all of the variables in the data file, creating a new data set. Output includes the correlation coefficient, t value for the correlation, degrees of freedom, and probability value. The user may also view a graphic scatterplot of the data to visually assess its linearity.

Test Statistics

This program performs a variety of statistical tests and analyses useful for specialized needs as well as for follow-up to other programs in Research Assistant. User options include a test to determine the significance of a difference between a sample mean and a population mean, tests of significance between variances of two related or independent samples, a test of the significance of two independent correlations, tests to determine the significance of differences between observed and expected proportions, a Scheffé test, and probability values for user specified z , t , and F values.

t tests and One-Way ANOVA

This program calculates independent and related t tests between two variables, and a one-way analysis of variance between 2 and 9 variables. Printed output includes sums of squares, mean squares, degrees of freedom, t or F values, and probability values.

Chi Square

When user data consists of frequencies in discrete categories that can be put into the form of a contingency table, the chi-square program can be used to determine the significance of difference between two independent variables and the degree of relationship between those two variables. Column variables (horizontal) may have 2-9 categories, and row variables (vertical) may have 2-7 categories; thus a 9×7 contingency table is accommo-

dated. Output for this program includes the chi-square value, degrees of freedom, and the associated probability value. Yates's correction for continuity and phi coefficients are calculated for 2×2 contingency tables. Tests for goodness of fit are also calculated to test whether an obtained distribution of frequencies departs significantly from some standard frequency distribution (normal, binomial, etc.).

Multiple Regression

This program performs a multiple regression analysis and allows the user to test the significance of difference between two R^2 values. This program can accommodate 1-8 independent variables, each with a maximum of 1,000 subjects. Output includes the raw regression equation, multiple correlation, R^2 , F value, probability value, and variable means and standard deviations. Additional user options include a printed correlation matrix and a complete list of obtained, predicted, and residual scores, along with the standard error of estimate.

PROGRAM AVAILABILITY

Research Assistant (including specific instructions, capabilities, and limitations for each program) is available in object code form on disk by sending \$10 (U.S.) to cover postage and reproduction costs to Marley Watkins. Non-U.S. requests should include \$15 (U.S.) for costs and airmail delivery.

REFERENCES

- BRUNING, J. L., & KINTZ, B. L. (1977). *Computational handbook of statistics*. Glenview, IL: Scott, Foresman.
- BUTLER, D. L., & JONES, S. K. (1987). A comparison of inexpensive statistical packages for Apple II microcomputers. *Behavior Research Methods, Instruments, & Computers*, *19*, 99-103.
- FERGUSON, G. A. (1981). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- GUILFORD, J. P., & FRUCHTER, B. (1978). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- NIE, N., HULL, C., JENKINS, J., STEINBRENNER, K., & BENT, D. (1975). *Statistical Package for the Social Sciences* (2nd ed.). New York: McGraw-Hill.
- PITTINGER, D. J., & HODGE, M. H. (1986). Statistical programs for Commodore computers. *Behavior Research Methods, Instruments, & Computers*, *18*, 50-52.
- SAS INSTITUTE. (1985). *SAS user's guide* (5th ed.). Cary, NC: SAS Institute.
- SNEDECOR, G. W., & COCHRAN, W. (1967). *Statistical methods*. Ames, IA: Iowa State University Press.
- STANDING, L. (1985). Psychostats 2.1: A statistics package in OS-65U BASIC. *Behavior Research Methods, Instruments, & Computers*, *17*, 510.

(Manuscript received October 10, 1986;
revision accepted for publication April 28, 1987.)