References

College of Agriculture (1984). *University of Idaho Long-Range Plan: College of Agriculture.* Moscow, ID: University of Idaho, College of Agriculture.

Drueckhammer, D.C. (1985). *Follow-up of College of Agriculture Graduates at Oklahoma State University: 1979-1983.* Stillwater, OK: Oklahoma State University, Department of Agricultural Education.

Webb, E.S., Stoner, T.W., and Vaclacik, R. (1978). *Follow-up of Agricultural Education Graduates from Texas A & M University for the Years 1965-1971.* College Station, TX: Texas A & M University, Department of Agricultural Education.

A NOTE

Accuracy of Microcomputer Regression Software

Dale J. Menkhaus, John Hewlett and Glen D. Whipple

With microcomputers and accompanying software becoming more accessible, the frequency of use by faculty and students in colleges of agriculture has increased. In addition to course specific applications, increased availability of statistical software has prompted a substitution of microcomputers for mainframes for statistical analyses, both for classroom exercises and research.

Objectives

The overall objective of this note is to document that it may be important for users of microcomputer statistical software, specifically of regression, to be aware of the computational accuracy of these programs. Since there are several regression packages available for microcomputers, it is impossible to check every program for its accuracy in this note. Thus, a simple procedure for testing the accuracy of regression programs is outlined and demonstrated using three anonymous routines. The purpose here is not to make recommendations with respect to specific regression packages; that decision is reserved to the individual user. The primary concern addressed in this note is with regard to computational accuracy which should provide input for selecting a regression package for use in the classroom and/or research.

Procedures

The procedure employed in this paper is that used by Wampler and demonstrated by Boehm, et al. Two problems defined by the following equations were used for the test. Values of the dependent variables (Y_1 and Y_2) for the test were calculated from the following equations.

$$Y_1 = 1 + 1X + 1X^2 + 1X^3 + 1X^4 + 1X^5$$

$$Y_2 = 1 + 0.1X + 0.01X^2 + 0.001X^3 + 0.0001X^4 + 0.00001X^5$$

Both equations are fifth degree polynominals. The values of the variable X were the integers, 0, 1, 2 —, 20. True values for the parameters are, of course, the values used to calculate the Y's, i.e., 1, 1, 1, 1, 1, and 1 for Y_1 and 1, 0.1, 0.01, 0.001, 0.0001, and 0.00001 for Y_2. There are no error terms incorporated into the equations and therefore the $R^2 = 1$ for each equation.

Simple correlation coefficients among the independent variables were all greater than 0.816, six of the ten were greater than 0.958, and three were greater than 0.986. The high linear association between the regressors and the large variation in the data partially explain why consistently accurate parameter estimates for the above equations are difficult to obtain (Boehm, et al., p. 757).

The two test problems have been classified by Wampler as being highly "ill-conditioned," with equation Y_1 slightly more ill-conditioned than the Y_2 equation.[1] Suffice it to say that the test problems are difficult to estimate. If computer routines successfully handle these problems, computational accuracy should not be a serious issue for less ill-conditioned cases. Five regression routines are reported. The regression packages were tested using an IBM-PC compatible microcomputer.

Results and Discussion

The computational accuracies of the regression routines tested were varied (Table 1). In most cases the estimated regression coefficients (B's) were reasonably accurate, with the exception of the estimate of B_0 for Y_1 from routine 2. The R^2 value is reported correctly in each of the routines for each equation estimated. As expected, the overall results tend to be better for equation Y_2 as compared to equation Y_1. This is due to the slightly more ill-conditioned nature of equation Y_1.

Estimates of the coefficient standard errors and the standard error of regression exhibit the greatest variation (Table 1). Each of these estimates should be equal to zero. While most estimates of the coefficient standard errors are close to zero, some were larger than their corresponding coefficient estimates, notably from routines 2 and 3 for equation Y_1. Thus, the B's using the classical t-test would be incorrectly judged non-significant. Routines 2 and 3 incorrectly estimated the standard error of regression for the Y_1 equation and, for some reason S^2 was estimated to be negative using routine 2.

Summary and Concluding Remarks

The primary purpose of this note was to document that it is important to check the computational accuracy of microcomputer regression software. A simple procedure for testing the accuracy of regression programs is provided and illustrated.

The results reported in this note suggest that there is some variation in computational accuracy among

Menkhaus, Hewlett and Whipple are research associate and associate professor, respectively, Department of Agricultural Economics, University of Wyoming, Laramie, WY 82071.

[1]The concept of an "ill-conditioned matrix" focuses on the expected severity of round-off errors generated in inversion. Several numbers have been proposed to measure the degree of ill-conditioning; however, empirical results have shown them to be inadequate (Ling). Newman discusses a commonly used measure, the P-condition.

Table 1. Summary of Estimates for Equations Y_1 and Y_2.

Equation and Routine	β_0	β_1	β_2	β_3	β_4	β_5	R^2	S[a/]	S^2
	Estimated Values of Regression Coefficients								
Routine 1									
Y_1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.08E-08	--
	(9.879E-09)[b/]	(1.085E-08)	(3.577E-09)	(4.657E-10)	(2.591E-11)	(5.155E-13)			
Y_2	1.0000	0.0999998	0.0100001	0.0010000	0.0001000	1.000E-05	1.0000	5.45E-07	--
Routine 2									
Y_1	0.484264	0.996926	1.00252	0.999991	1.00001	1.0000	1.0000	1197.43	-1433830
	--	(1199.15)	(395.402)	(51.4839)	(2.86434)	(0.0569883)			
Y_2	0.999998	0.10000	0.010000	1.0000E-03	1.0000E-04	9.999998E-06	1.0000	0.0106192	-1.12767E-04
	--	(0.0106345)	(0.00350656)	(4.565764E-04)	(2.540197E-05)	(5.053917E-07)			
Routine 3									
Y_1	1.000011	0.999999	1.0000	0.999999	1.0000	1.0000	1.0000	4.292818	--
	--	(4.299014)	(1.417531)	(0.184571)	(0.010268)	(0.000204)			
Y_2	1.0000	0.1000	0.01000	0.0010	0.0001	0.00001	1.0000	0.000081	--
	--	(0.000081)	(0.000026)	(0.000003)	(0.0000)	(0.0000)			

a/ The true values for S and $S^2 = 0$.
b/ Estimated coefficient standard errors are reported in parentheses.

microcomputer regression routines, particularly regarding estimates of coefficient standard errors and the standard error of regression for the test problems. Such results demonstrate the importance of checking the computational accuracy of statistical packages, and for that matter all software, before using them in the classroom and/or in research.

Furthermore, checking the accuracy of new software should become a routine practice of users. The method explained here can be used to check the accuracy of regression softwares. Other microcomputer statistical software and other packages, e.g., linear programming algorithms, can be checked against those available from campus mainframes. While it has not always been the case that mainframe software packages have been computationally accurate, most "big name" packages are now accurate. In fact, the two programs which proved to be most accurate in this study are among two well-known statistical packages for mainframes.

Additional Observations

In addition to computational accuracy, there is, of course, noticeable variation in other characteristics of regression software. Depending on how the regression package is to be used, some of these characteristics may be important. These might include

— Easy data entering and transforming

— Availability of graphical and analysis or plotting

— Easy data exporting to or importing from other regression packages

— Regression results Formatted for easy reading of regression results

— Regression results complete, e.g., the Durbin-Watson statistic, adjusted R^2, etc.

— A complete package in terms of the types of analyses available.

Finally, no matter how many "whistles and bells" are available in the regression software package, the routine is of little value if it is not computationally accurate.

References

Boehm, W.T., D.J. Menkhaus, and J.B. Penn. November 1986. "Accuracy of Least Squares Computer Programs: Another Reminder." *Amer. J. of Agr. Econ.*, Vol. 58, No. 4, pp. 757-760.

Ling, Robert F. August 1974. "Least Squares Computation, Ill-Conditioning, and SWEEP." Paper presented at the American Statistical Association Meeting, St. Louis, Missouri.

Newman, Morris. 1962. "Matrix Computations." *Survey of Numerical Analysis*, ed. John Todd, pp. 239-43. New York: McGraw-Hill Book Co.

Wampler, R.H. 1970. "A Report on the Accuracy of Some Widely Used Least Squares Computer Programs." *J. Amer. Statis. Assoc.*, Vol. 65, pp. 549-565.