Volume 13, Number 1 Spring 1984

A Comparative Analysis of Books on Multiple Regression

Dennis W. Leitner

and

Randall E. Schumacker

Southern Illinois University at Carbondale

ABSTRACT

The purpose of this article was to compare 33 books on multiple regression that have appeared in past years. The books were compared on topics covered; year and publisher; orientation; level of presentation; background required; published reviews; and readability. This information should aid students and instructors in selecting a multiple regression book for a given topic, approach, or need.

INTRODUCTION

In the past decade and a half, many books and reference works on the topic of multiple regression have become available. This is probably due to a combination of several factors, some of which might be the following: (1) The widespread availability of computers and computer programs. (2) The pioneering work of Robert Bottenberg, Joe Ward, Earl Jennings and their disciples from Texas, and (3) The excellent articles by Jacob Cohen entitled "Multiple regression as a general data-analytic device" and Richard Darlington entitled "Multiple regression in psychological and research practice", both of which appeared in Psychological Bulletin. The books were compared on the following information: (1) topics covered; (2) year and publisher; (3) orientation (theoretical vs applied); (4) level of presentation (textbook vs reference); (5) background required (basic statistics vs matrix algebra); (6) reference list of published reviews; and (7) readability.

Ellinger (1981) presented a historical overview of readability formulae from 1921 to the present. She included the criteria on which they were based, factors evaluated, reliability and validity. The authors selected the Flesch Ease Formula (Flesch, 1948) because it is appropriate for adult reading material and does not require use of a word list. The Flesch Ease Formula is:

$$\text{Score} = 206.835 - (.846 * \text{NSYL}) - (1.015 * \text{SL})$$

where: NSYL = number of syllables per a 100 word sample

SL = average sentence length

The reading ease score can range from 0 to 100 with a lower score indicating a more difficult reading level (see Table 3).

METHOD

Book Selection

The books selected for comparison represent those from the social sciences and statistical literature that an educational researcher might be inclined to use. The list is by no means exhaustive. The books are listed alphabetically by author in a separate section of the references.

Criteria for Topic Inclusion

The books were compared on selected topics that the authors felt readers would be interested in. A particular topic was indicated as being covered in a given book if sufficient presentation or explanation was present, e.g. an entry in the table of contents, subject index, or several pages of discussion. For example, linear regression meant the straight-line one predictor variable regression model.

Anova compared to regression implied a comparison of the two as special cases of the general linear model. Matrix algebra referred to whether or not the book contained matrix algebra computations or had an appendix with matrix algebra rules and procedures. The covariance topic included the discussion of analysis of covariance. The other selected topics are self-explanatory.

RESULTS

Table 1 indicates a comparison of topics among the books. Most books covered the following topics: linear regression, curvilinear, polynomial or nonlinear regression, zero-order correlation, matrix algebra, partial/semi-partial correlation, and anova compared to regression, respectively.

Several books published between 1966 and 1976 had included computer programs while later books did not because of the popularity of several statistical packages. For example, Younger (1979) published a book which included comparisons among SAS, SPSS, and BMDP statistical packages in performing numerous applied regression examples.

Table 2 indicates a comparison among the books according to year, publisher, orientation, presentation, and background. Most books appeared in the 1970's (1960's, n=9; 1970's, n=18; 1980's, n=6). The majority had an applied orientation (applied, n=21; theoretical, n=9; both, n=3) with both a textbook and reference level of presentation (textbook, n=9; reference, n=12; both, n=12). Most books also required a basic statistics background (basic statistics, n=21; matrix algebra, n=12). Books with a theoretical orientation usually required knowledge of matrix algebra.

Published reviews were found for many of the books. These are listed alphabetically by author in a separate section of the references. The reviews permit an individual to read about another person's opinion of a book the authors have reviewed.

Selected Topics by Author

Author Name	Book Topics											
	a	b	c	d	e	f	g	h	i	j	k	l
Belsley	x	x		x		x	x	x	x	x	x	x
Chatterjee	x	x					x	x		x	x	
Cohen (1)	x	x	x	x	x	x	x	x	x	x	x	x
Cohen (2)	x	x	x	x	x	x	x	x	x	x	x	x
Draper (1)	x	x	x		x	x		x	x	x	x	
Draper (2)	x	x	x		x	x	x	x	x	x	x	
Dunn	x	x	x		x	x						x
Edwards (1)	x	x			x	x						
Edwards (2)	x	x	x		x	x	x	x	x	x		x
Fraser	x			x					x			
Freund	x	x	x		x	x	x		x	x		x
Goldberger	x	x			x	x	x		x	x		
Graybill (1)	x	x	x		x				x			x
Graybill (2)	x	x	x	x	x	x			x	x	x	
Gunst	x	x			x		x	x	x	x	x	
Haitovsky	x								x			x
Huang	x	x		x	x		x	x	x	x		
Kelly	x	x	x	x	x	x						x
Kerlinger	x	x	x	x	x	x	x	x	x		x	x
Kleinbaum	x	x	x	x	x	x	x	x	x	x	x	x
Koerts	x				x				x	x		
Lewis		x	x			x		x				x
McNeil	x	x	x	x	x	x		x			x	x
Pedhazur	x	x	x	x	x	x	x	x	x	x	x	x
Plackett	x	x	x						x	x		
Rao	x		x		x	x			x			
Searle	x		x					x	x	x	x	x
Smillie	x	x			x			x	x	x	x	
Sprent	x	x	x		x	x			x	x		x
Ward	x	x			x			x	x			x
Williams	x	x	x		x	x		x			x	x
Wonnacott	x	x	x		x	x	x	x	x	x	x	x
Younger	x	x		x	x	x	x	x	x	x	x	

KEY: a = linear regression
b = curvilinear, polynomial, or nonlinear regression
c = anova compared to regression
d = multivariate techniques
e = zero-order correlation
f = partial/semi-partial correlation
g = multicollinearity
h = dummy, effect and/or contrast coding
i = matrix algebra
j = residual analysis/outliers
k = variable selection methods
l = covariance

Table 2

Comparative Book Information

	Author	Year	Publisher	Orientation a	Orientation b	Presentation c	Presentation d	Background e	Background f
	Belsley	1980	John Wiley	x	x	x		x	
	Chatterjee	1977	John Wiley	x			x		x
	Cohen (1)	1975	John Wiley	x		x	x		x
	Cohen (2)	1983	John Wiley	x		x	x		x
→	Draper (1)	1966	John Wiley	x		x	x		x
	Draper (2)	1981	John Wiley	x		x	x	x	
	Dunn	1974	John Wiley	x			x		x
	Edwards (1)	1976	W.H.Freeman	x		x			x
	Edwards (2)	1979	W.H.Freeman	x		x			x
	Fraser	1979	McGraw-Hill		x		x	x	
	Freund	1979	Marc-Dekker	x			x		x
	Goldberger	1968	MacMillan		x		x	x	
	Graybill (1)	1961	McGraw-Hill	x	x	x	x	x	
	Graybill (2)	1976	Duxbury		x		x	x	
	Gunst	1980	Marc-Dekker	x		x	x		x
	Haitovsky	1973	Hafner		x		x	x	
	Huang	1970	John Wiley		x		x	x	
	Kelly	1969	SIU press	x		x			x
	Kerlinger	1973	Holt,R & W	x		x	x		x
	Kleinbaum	1978	Duxbury	x		x	x		x
	Koerts	1969	Rotterdam	x	x	x		x	
	Lewis	1978	SIU press	x			x		x
→	McNeil	1975	SIU press	x		x			x
→	Pedhazur	1982	Holt,R & W	x		x	x		x
	Plackett	1960	Oxford press		x		x	x	
	Rao	1965	John Wiley		x		x	x	
	Searle	1971	John Wiley		x		x	x	
	Smillie	1966	Ryerson	x		x			x
	Sprent	1969	Methuen		x	x	x		x
→	Ward	1973	Prentice	x		x	x		x
	Williams	1974	MSS corp.	x		x			x
	Wonnacott	1981	John Wiley	x		x			x
	Younger	1979	Duxbury	x		x	x		x

KEY: a = applied
b = theoretical
c = textbook
d = reference
e = matrix algebra
f = basic statistics

Table 3 indicates the Flesch Reading Ease Scale used to interpret the book score listed in Table 4. All books ranged from fairly difficult to very difficult which would be expected given the topics discussed. The readability measure does not take into consideration the numerous formulae, graphs, notation and mathematics. It does however provide some indication of readability for comparison among the books as well as a general indication of reading complexity compared to other types of reading material.

CONCLUSION

The information provided permits comparisons among several books of multiple regression published over past years. Certain topics were indicated as appearing in the majority of the books. Most of the books reviewed emphasized an applied orientation with a basic statistics background requirement. Additional inquiry about certain books is possible by referring to the published reviews. The Flesch Ease Formula was used to compute a score on each book. The books reflected a difficult reading level comparable to scientific and academic text.

Most books had an outstanding feature which became apparent during the review process. For example, Belsley covered analysis of outliers and sources of multicollinearity. Chatterjee covered multicollinearity, autocorrelation and ridge regression extremely well. Cohen (1) and Cohen (2) had the widest range of topics covered and included one of the few discussions of power. Draper (1), Draper (2), and Gunst present the analysis of residuals/outliers and variable selection techniques the best. Edwards (1) and Edwards (2) afford an excellent introduction to linear regression with the presentation of different designs for analysis with dummy, effect, and contrast coding. Graybill (1) and Graybill (2) offer a broad coverage of topics at an advanced level using a matrix algebra

Table 3

Flesch Reading Ease Scale[a]

Reading Ease Scale	Description	Typical Magazine	Grade Level
0 to 30	Very Difficult	Scientific	
30 to 50	Difficult	Academic	
50 to 60	Fairly Difficult	Quality	
60 to 70	Standard	Digests	8th
70 to 80	Fairly Easy	Slick-fiction	7th
80 to 90	Easy	Pulp-fiction	6th
90 to 100	Very Easy	Comics	5th
100			4th

[a]Adapted from Flesch, 1948, p 230

Table 4

Readability Comparisons

R.E. Score	Author	Description
16.5	Cohen (2)	
18.1	Pedhazur	Very Difficult
23.1	Belsley	
32.6	Kleinbaum	
32.9	Haitovsky	
33.1	Gunst	
33.8	Huang	
35.1	Cohen (1)	
35.9	Goldberger	
35.9	Smillie	
36.8	Draper (2)	
38.2	Fraser	Difficult
39.5	Kerlinger	
42.7	Rao	
45.4	Sprent	
46.2	Wonnacott	
46.5	Koerts	
48.5	Dunn	
48.8	Ward	
50.7	Draper (1)	
52.0	Chatterjee	
52.7	Lewis	
53.1	Searle	
54.4	Graybill (2)	
54.9	Freund	
55.9	Edwards (1)	Fairly Difficult
56.6	McNeil	
56.7	Williams	
56.9	Kelly	
57.2	Younger	
57.2	Graybill (1)	
57.4	Plackett	
60.0	Edwards (2)	

approach. Kelly, Lewis, and McNeil propose model formulation to test given research hypotheses. Kerlinger and Pedhazur cover dummy, effect, and contrast coding well. Pedhazur additionally included a computer program on LISREL. Kleinbaum provides a broad coverage of all topics with excellent multivariate examples. Williams provides excellent examples on coding repeated measure designs. And finally, Younger provides computer applications using SAS, BMDP, and SPSS. Overall, selection of a specific book for classroom use is in the "eyes of the beholder", but this information should permit an alternative to experimentation or chance selection.

REFERENCES

Books Reviewed

Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons.

Chatterjee, S., & Price, B. (1977). Regression Analysis by Example. New York: John Wiley & Sons.

Cohen, J., & Cohen, P. (1975). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. New York: John Wiley & Sons.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd Ed.). New York: John Wiley & Sons.

Draper, N. R., & Smith, H. (1966). Applied Regression Analysis. New York: John Wiley & Sons.

Draper, N. R., & Smith, H. (1981). Applied Regression Analysis (2nd Ed.). New York: John Wiley & Sons.

Dunn, O. J., & Clark, V. A. (1974). Applied Statistics: Analysis of Variance and Regression. New York: John Wiley & Sons.

Edwards, A. L. (1979). An Introduction to Linear Regression and Correlation. San Francisco: W. H. Freeman.

Edwards, A. L. (1979). Multiple Regression and the Analysis of Variance and Covariance. San Francisco: W. H. Freeman.

Fraser, D. A. S. (1979). Inference and Linear Models. Great Britian: McGraw-Hill.

Freund, R. J., & Minton, P.O. (1979). Regression Methods. A Tool for Data Analysis. New York: Marcel Dekker.

Goldberger, A. S. (1968). Topics in Regression Analysis. New York: MacMillan.

Graybill, Franklin A. (1961). An Introduction to Linear Statistical Models (Vol. 1). New York: McGraw-Hill.

Graybill, Franklin A. (1976). Theory and Applications of the Linear Model. Massachusetts: Duxbury Press.

Gunst, R. F. & Mason, R. L. (1980). Regression Analysis and its Application: A Data-oriented Approach. New York: Marcel Dekker.

Haitovsky, Y. (1973). Regression Estimation from Grouped Observations. New York: Hafner Press.

Huang, D. S. (1970). Regression and Econometric Methods. New York: John Wiley & Sons.

Kelly, F. J., Beggs, D. L., & McNeil, K. A. (with Eichelberger, T. and Lyon, J.) (1969). Multiple Regression Approach: Research Design in the Behavioral Sciences. Illinois: Southern Illinois University Press.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple Regression in Behavioral Research. New York: Holt, Rinehart & Winston.

Kleinbaum, D. G., & Kupper, L. L. (1978). Applied Regression Analysis and other Multivariable Methods. California: Duxbury Press.

Koerts, J., & Abrahamse, A. P. J. (1969). On the Theory and Application of the General Linear Models. Netherlands: Rotterdam University Press.

Lewis, E. L., & Mouw, J. T. (1978). The Use of Contrast Coefficients (Supplement to McNeil, Kelly, and McNeil, Testing Research Hypotheses Using Multiple Linear Regression). Illinois: Southern Illinois University Press.

McNeil, K. A., Kelly, F. J., & McNeil, J. T. (1975). Testing Research Hypotheses Using Multiple Linear Regression. Illinois: Southern Illinois University Press.

Pedhazur, E. J. (1982). Multiple Regression in Behavioral Research Explanation and Prediction (2nd Ed.). New York: Holt, Rinehart & Winston.

Plackett, R. L. (1960). Principles of Regression Analysis. London: Oxford University Press.

Rao, C. R. (1973). Linear Statistical Inference and its Application (2nd Ed.). New York: John Wiley & Sons.

Searle, S. R. (1971). Linear Models. New York: John Wiley & Sons.

Smillie, K. W. (1966). An Introduction to Regression and Correlation. Toronto: Ryerson Press.

Sprent, P. (1969). Models in Regression and Related Topics. London: Methuen.

Ward Jr., J. H., & Jennings, E. (1973). Introduction to Linear Models. New Jersey: Prentice-Hall.

Williams, J. D. (1974). Regression Analysis in Educational Research. New York: MSS Information Corporation.

Wonnacott, T. H., & Wonnacott, R. J. (1981). Regression: A Second Course in Statistics. New York: John Wiley & Sons.

Younger, M. (1979). A Handbook for Linear Regression. California: Duxbury Press.

Belsley

Patterson, H. D. (1981). Biometrics. 37. 862-863.

Chatterjee

Bargmann, R. E. (1979). Contemporary Psychology. 24 (1). 45-46.

Cohen (1)

Tufte, E. R. (1979). Journal of the American Statistical Assn. 74. 935.

Henry, Neil W. (1977). Contemporary Psychology. 6. 320.

Cohen (2): none found

Draper (1)

Mallows, C. L. (1967). American Statistical Association Journal. 62. 1075.

Draper (2): none found

Dunn

Imrey, P. B. (1976). Journal of American Statistical Association. 71. 245.

(1974). Choice. 7. 1170.

Edwards (1)

Timm, N. H. (1977). Contemporary Psychology. 22. 174.

Edwards (2)

Timm, N. H. (1980). Choice. 16. 1614.

Fraser: none found

Freund

Dwass, M. (1980). Science Books and Films. 16. 12.

Goldberger

Mandansky, A. (1970). American Statistical Association Journal. 64. 1965.

Graybill (1)

Cochran, W. G. (1962). American Statistical Association Journal. 57. 503.

Graybill (2)

Sen, P. K. (1978). American Statistical Association Journal. 73. 221.

Gunst

McGuigan, R. (1979). Science Books & Films. 15 (1). 194.

Haitovsky: none found

Huang

Geraci, V. J. (1972). American Statistical Association Journal. 67. 961.

Kelly

Wasik, J. L. (1971). Educational and Psychological Measurement. 31. 575-577.

Kerlinger

Elashoff, J. D. (1975). American Statistical Association Journal. 70 959.

(1974). Choice. 10. 1753.

Walsh, J. A. (1974). Educational and Psychological Measurement. 34. 724-728.

Kleinbaum

Bhattacharya, H. T. (1979). American Statistical Association Journal. 74. 732.

Koerts: none found

Lewis: none found

McNeil

Burstein, L. (1976). American Educational Research Association Journal. 13. 196-301.

Pedhazur: none found

Plackett

Cochran, W. G. (1961). American Statistical Association Journal. 56. 418.

Rao

Efron, B. (1967). American Statistical Association Journal. 62. 29

Searle

John, P. W. M. (1972). Journal of the American Statistical Association. 67. 954.

Smillie: none found

Sprent

(1970). Choice. 7. 717.

Ward

Burstein, L. (1976). American Education Research Journal. 13. 296-301.

Williams: none found

Wonnacott: none found

Younger

Askin, R. G. (1980). Journal of Quality Technology. 12. 55-56.

(1980). Choice. 17. 111.

Readability

Ellinger, P. J. (1981). Use of Readability Formulas in the Formative Evaluation of Patient Education Materials. Unpublished master's thesis, Southern Illinois University. Carbondale, Illinois.

Flesch, R. F. (1948) A new readability yardstick. Journal of Applied Psychology. 32. 221-233.

Articles

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.