Linear Regression Analysis

G.A.F. SEBER

A VOLUME IN THE WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS: RALPH A. BRADLEY, J. STUART HUNTER, DAVID G. KENDALL, & GEOFFREY S. WATSON—ADVISORY EDITORS

LINEAR REGRESSION ANALYSIS

G.A.F. SEBER

Professor of Statistics
University of Auckland
Auckland, New Zealand

John Wiley & Sons

New York · London · Sydney · Toronto

Library of Congress Cataloging in Publication Data:

Seber, George Arthur Frederick.
Linear regression analysis.

(Wiley series in probability and mathematical statistics)
Includes bibliographical references and index.
1. Regression analysis. I. Title.
QA278.2.S4 519.5'36 76-40117
ISBN 0-471-01967-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Linear Regression Analysis

Preface

Regression analysis is an often used tool in the statistician's toolbox. The theory is elegant and the computational problems intriguing, so that both "pure" and "applied" statisticians can feel at home in the subject. For example, among the theoreticians there is still an ongoing interest in least squares with all its generalizations and special cases. At the same time practitioners continue to develop a wide range of graphic methods for testing models and examining underlying assumptions. As numerical analysis and statistics have slowly intertwined, statisticians have been made to realize the difficulties associated with certain time-honored computational procedures. The development of regression computer programs that are efficient and accurate is now recognized as an important part of statistical research.

However, this continuing research interest in regression across the pure–applied spectrum creates problems for the textbook writer. As demonstrated by current books on the subject, the material can be presented at a variety of levels of mathematical difficulty ranging from the very general treatments of Seber [1966], Searle [1971], and Rao [1973], for example, to the more discursive approaches of Williams [1959] and Sprent [1969]. Clearly such a variety of books is needed because there is a wide diversity of readers. However, while teaching the subject continuously over the past 10 years I have become more and more aware of the need for a suitable university text that takes a middle road between giving no proofs of results and giving proofs in complete generality. Because regression analysis and much of analysis of variance are concerned with full rank models, it would appear that the less than full rank case tends to be overemphasized for the sake of greater generality. In particular, the generalized inverse has been rather overworked to the detriment of the simple geometric ideas that underlie least squares. Clearly the generalized inverse has its uses, but its role needs to be kept in perspective.

Regression analysis is an applied subject describing methods for handling data; ideally, any theoretical course should be backed up by practical work. This raises the question of whether a single textbook should try to deal, in detail, with both the theoretical and computational aspects of the subject. Clearly such a task is not easy and in this age of statistical package programs it would seem more appropriate to treat the two aspects quite separately. For example, procedures that have an elegant theory may be computationally unsatisfactory, whereas messy, complicated algorithms may be efficient and accurate. Because packages vary from place to place according to the local computer facilities, the ultimate solution might be to have a theoretical textbook that includes the computational aspects of regression in broad outline only and a practical "manual" giving numerical examples and details of packages; a forerunner of the latter type of book is Daniel and Wood [1971].

With the above thoughts in mind I have endeavored to write a theoretical book that is satisfying for the mathematically minded reader but that does not lose the reader in generalities. I have also endeavored to give an up-to-date account of computational methods and algorithms currently in use without getting entrenched in minor computing details. Since the research literature on regression continues to grow rapidly, I have surveyed the better-known statistical journals with the aim of producing a book that I hope will be useful as a general reference. The basic prerequisites for reading this book are a good knowledge of matrix algebra and some acquaintance with straight-line regression and simple analysis of variance models.

The first four chapters provide a fairly standard formal treatment of least squares fitting and hypothesis testing for the multiple linear regression model. In Chapter 1 expectation and covariance operators for vectors of random variables are introduced gently, and in Chapter 2 the multivariate normal distribution and certain theorems on quadratic forms are considered. Chapter 3 deals with least squares estimation and includes generalized least squares, the less than full rank case, and estimation with restrictions. Chapter 4 considers, in detail, the F-test for a linear hypothesis, and in Chapter 5 there is a discussion of confidence intervals and simultaneous inference as applied to regression models; prediction and inverse prediction (discrimination) intervals are also considered. Chapter 6 examines the assumptions underlying the least squares theory and various methods of testing these assumptions are provided. Because straight-line and polynomial fitting are important techniques, Chapters 7 and 8 are devoted to these two topics, respectively. Chapter 9 exploits the close relationship that exists between regression and analysis of variance models and provides simple procedures for carrying out an analysis of variance; attention is confined mainly to balanced (orthogonal) designs. In Chapter 10 the analysis of covariance is also considered from a regression viewpoint and, because of close ties with analysis of covariance, the topic of missing observations is considered in detail. Chapters 11 and 12 deal with the computational aspects of regression analysis: Chapter 11 is devoted to algorithms for least squares fitting and Chapter 12 considers the problem of choosing the best regression subset from a set of likely regressor (independent) variables.

Appendixes A and B contain a number of matrix results whose proofs are not always readily accessible, and Appendix C describes probability plotting. Appendixes D, E, and F give some statistical tables which are useful in simultaneous inference. Finally, there is a set of outline solutions for the exercises.

It has not been easy to find or make up theoretical problems that are relevant and yet not too difficult. It is hoped that the 200 or so problems scattered throughout the book will not only help the student but also provide some ideas for teachers.

This book is based on several courses that I have been giving at Auckland University, New Zealand, during the past 10 years and I wish to thank the many students who have stimulated my teaching interest in the subject. Special thanks are also due to Heather Lucas for reading a first draft and to Peggy Haworth for typing most of the manuscript.

ACKNOWLEDGMENTS

For permission to reproduce certain published tables and figures, thanks are due to the authors and Editors of *Biometrika* (Appendix E), *Journal of the American Statistical Association* (Tables 5.1, 5.2, and Appendix D), *Journal of the Royal Statistical Society, Series B* (Appendix F), and *Technometrics* (Table 5.3).

G. A. F. SEBER

Auckland, New Zealand
July 1976

Contents

CHAPTER 1

Vectors of Random Variables

1.1 NOTATION

Matrices and vectors are denoted by boldface letters $\mathbf{A}$ and $\mathbf{a}$, respectively, and scalars by italics. Random variables are represented by capital letters, and their values by lowercase letters (for example, Y and y, respectively). This use of capitals for random variables, which seems to be widely accepted, is particularly useful in regression when distinguishing between fixed and random regressor (independent) variables. However, it does cause problems because a vector of random variables, $\mathbf{Y}$, say, then looks like a matrix. Occasionally in Chapter 11, because of a shortage of letters, a boldface lowercase letter represents a vector of random variables.

If X and Y are random variables then the symbols $E[Y]$, var$[Y]$, cov$[X, Y]$, and $E[X|Y=y]$ (or, more briefly, $E[X|Y]$) represent expectation, variance, covariance, and conditional expectation, respectively.

The $n \times n$ matrix with diagonal elements $d_1, d_2, \ldots, d_n$ and zeros elsewhere is denoted by diag$(d_1, d_2, \ldots, d_n)$, and when all the d_i's are unity we have the identity matrix $\mathbf{I}_n$.

If $\mathbf{a}$ is an $n \times 1$ column vector with elements $a_1, a_2, \ldots, a_n$, we write $\mathbf{a} = [(a_i)]$, and the *length* or *norm* of $\mathbf{a}$ is denoted by $\|\mathbf{a}\|$. Thus

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}'\mathbf{a}} = \left(a_1^2 + a_2^2 + \cdots + a_n^2\right)^{1/2},$$

The vector with elements all equal to unity is represented by $\mathbf{1}_n$.

If the $m \times n$ matrix $\mathbf{A}$ has elements a_{ij} we write $\mathbf{A} = [(a_{ij})]$, and the sum of the diagonal elements, called the trace of $\mathbf{A}$, is denoted by tr$\,\mathbf{A}$ $(= a_{11} + a_{22} + \cdots + a_{kk}$, where k is the smaller of m and n). The transpose of $\mathbf{A}$ is represented by $\mathbf{A}' = [(a'_{ij})]$, where $a'_{ij} = a_{ji}$. If $\mathbf{A}$ is square its determinant is written $|\mathbf{A}|$, and if $\mathbf{A}$ is nonsingular its inverse is denoted by $\mathbf{A}^{-1}$. The space