

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

Copyright © 1977 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means, nor transmitted, nor translated into a machine language without the written permission of the publisher.

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

- 1 VECTORS OF RANDOM VARIABLES 1
 - 1.1 Notation 1
 - 1.2 Linear Regression Models 2
 - 1.3 Expectation and Covariance Operators 8
 - Exercises 1a, 12
 - 1.4 Mean and Variance of Quadratic Forms 13
 - Exercises 1b, 16
 - 1.5 Independence of Random Variables 17
 - Exercises 1c, 19
 - 1.6 Chi-Squared Distribution 19
 - Miscellaneous Exercises 1, 20

- 2 MULTIVARIATE NORMAL DISTRIBUTION 22
 - 2.1 Definition 22
 - Exercises 2a, 27
 - 2.2 Moment Generating Function 27
 - Exercises 2b, 31
 - 2.3 Independence of Normal Variables 32
 - Exercise 2c, 35
 - 2.4 Quadratic Forms in Normal Variables 36
 - Exercises 2d, 40
 - Miscellaneous Exercises 2, 40

- 3 LINEAR REGRESSION: ESTIMATION AND DISTRIBUTION THEORY . . . 42
 - 3.1 Least Squares Estimation 42
 - Exercises 3a, 47

3.2	Properties of Least Squares Estimates	48
	Exercises 3b, 50	
3.3	Estimation of σ^2	51
	Exercises 3c, 54	
3.4	Distribution Theory	54
	Exercises 3d, 56	
3.5	Orthogonal Structure in the Design Matrix	56
	Exercises 3e, 59	
3.6	Generalized Least Squares	60
	Exercises 3f, 64	
3.7	Introducing Further Regressors	65
	3.7.1 <i>General theory</i> , 65	
	3.7.2 <i>One extra variable</i> , 68	
	3.7.3 <i>Two-step least squares</i> , 69	
	3.7.4 <i>Two-step residuals</i> , 70	
	Exercises 3g, 71	
3.8	Design Matrix of Less than Full Rank	72
	3.8.1 <i>Least squares estimation</i> , 72	
	a. Reduce the model to one of full rank, 73	
	b. Impose identifiability constraints, 74	
	c. Compute a generalized inverse, 76	
	Exercises 3h, 78	
	3.8.2 <i>Estimable functions</i> , 79	
	Exercises 3i, 81	
	3.8.3 <i>Introducing further regressors</i> , 82	
	Exercises 3j, 84	
3.9	Estimation with Linear Restrictions	84
	3.9.1 <i>Method of Lagrange multipliers</i> , 84	
	3.9.2 <i>Method of orthogonal projections</i> , 86	
	Exercises 3k, 87	
3.10	Other Methods of Estimation	88
	3.10.1 <i>Biased estimation</i> , 88	
	3.10.2 <i>Nonnegative estimation</i> , 90	
	3.10.3 <i>Censored data</i> , 90	
	3.10.4 <i>Robust estimation</i> , 91	
3.11	Optimal Design	92
	Miscellaneous Exercises 3, 93	

4	LINEAR REGRESSION: HYPOTHESIS TESTING	96
4.1	The F -Test	96
4.1.1	<i>Derivation</i> , 96	
	Exercises 4a, 99	
4.1.2	<i>Motivation of the F-test</i> , 99	
4.1.3	<i>Some examples</i> , 101	
4.1.4	<i>The straight line</i> , 105	
4.1.5	<i>Significant F-test</i> , 108	
	Exercises 4b, 109	
4.2	Multiple Correlation Coefficient	110
	Exercises 4c, 115	
4.3	A Canonical Form For H	115
	Exercises 4d, 117	
4.4	Goodness of Fit Test	117
4.5	Design Matrix of Less than Full Rank	120
4.5.1	<i>The F-test and projection matrices</i> , 120	
4.5.2	<i>Testable hypotheses</i> , 121	
	Exercises 4e, 122	
4.6	Hypothesis Testing with Initial Constraints	123
	Miscellaneous Exercises 4, 123	
5	CONFIDENCE INTERVALS AND REGIONS	125
5.1	Simultaneous Interval Estimation	125
5.1.1	<i>The problem of simultaneous inferences</i> , 125	
	a. Bonferroni t -intervals, 126	
	b. Maximum modulus t -intervals, 127	
	c. Scheffé's S -method, 128	
5.1.2	<i>Comparison of methods</i> , 130	
5.1.3	<i>Hypothesis testing and confidence intervals</i> , 132	
5.1.4	<i>Confidence regions</i> , 133	
5.2	Confidence Bands for the Regression Surface	134
5.3	Prediction Intervals for the Response	136
5.4	Enlarging the Regression Matrix	137
	Miscellaneous Exercises 5, 139	
6	DEPARTURES FROM UNDERLYING ASSUMPTIONS	140
6.1	Bias	141

6.1.1	<i>Bias due to underfitting</i> , 141	
6.1.2	<i>Bias due to overfitting</i> , 143	
6.1.3	<i>Bias and random regressor variables</i> , 144	
6.2	Incorrect Dispersion Matrix	144
6.2.1	<i>General case</i> , 144	
6.2.2	<i>Diagonal dispersion matrix</i> , 146	
	a. Bartlett's test, 147	
	b. Cochran's test, 147	
	c. Hartley's F -max test, 148	
6.3	Robustness of the F -Test to Nonnormality	149
6.3.1	<i>Effect of the regressor variables</i> , 149	
6.3.2	<i>Quadratically balanced F-tests</i> , 151	
6.4	Regressor Variables Measured with Error	155
6.4.1	<i>Random errors</i> , 155	
	a. Bias, 156	
	b. Standard errors, 157	
6.4.2	<i>Roundoff errors</i> , 158	
6.4.3	<i>Some working rules</i> , 159	
6.5	Models with Random Regressors	160
6.6	Analysis of Residuals	162
6.6.1	<i>Definition and properties</i> , 162	
6.6.2	<i>Residual plots</i> , 163	
	a. Normal probability plot, 164	
	b. Plot residual versus fitted value, 165	
	c. Plot residual against an omitted factor (e.g., time), 166	
	d. Plot residual against each regressor, 169	
	e. Plot x_j against x_k , 170	
6.6.3	<i>Statistical tests based on residuals</i> , 170	
6.6.4	<i>Partial residual plots</i> , 170	
6.6.5	<i>Transformed residuals</i> , 172	
6.7	Transforming the Data	173
	Miscellaneous Exercises 6, 176	
7	STRAIGHT-LINE REGRESSION	177
7.1	Introduction	177
7.2	Confidence Intervals and Bands	179
7.2.1	<i>Confidence intervals for slope and intercept</i> , 179	
7.2.2	<i>Confidence interval for $-\beta_0/\beta_1$</i> , 180	

7.2.3	<i>Prediction intervals and bands</i> , 182	
7.2.4	<i>Prediction intervals for the response</i> , 186	
7.2.5	<i>Optimal allocation of observations</i> , 186	
7.2.6	<i>Inverse prediction (discrimination)</i> , 187	
	a. Alternative estimators, 189	
	b. Replicated observations, 190	
7.3	<i>Straight Line Through the Origin</i>	191
7.4	<i>Weighted Least Squares</i>	193
7.4.1	<i>Known weights</i> , 193	
7.4.2	<i>Unknown weights</i> , 194	
	a. Maximum likelihood method, 195	
	b. Least squares method, 197	
7.5	<i>Comparing Straight Lines</i>	197
7.5.1	<i>General model</i> , 197	
7.5.2	<i>Test for parallelism</i> , 198	
7.5.3	<i>Test for coincidence</i> , 200	
7.5.4	<i>Test for concurrence</i> , 201	
	a. <i>x</i> -coordinate known, 201	
	b. <i>x</i> -coordinate unknown, 203	
7.5.5	<i>Use of dummy categorical regressors</i> , 204	
7.5.6	<i>Test for equal variances</i> , 205	
7.6	<i>Two-Phase Linear Regression</i>	205
7.6.1	<i>Least squares estimation with known changeover point</i> , 206	
7.6.2	<i>Hypothesis test that changeover point is a given value</i> , 207	
7.6.3	<i>Unknown changeover point</i> , 208	
7.7	<i>Random Regressor Variable</i>	210
	<i>Miscellaneous Exercises 7</i> , 212	

8	POLYNOMIAL REGRESSION	214
8.1	Polynomials in One Variable	214
8.1.1	<i>The problem of ill-conditioning</i> , 214	
8.1.2	<i>Choosing the degree</i> , 217	
8.2	Orthogonal Polynomials	218
8.2.1	<i>General statistical properties</i> , 218	
8.2.2	<i>Generating orthogonal polynomials</i> , 220	
8.2.3	<i>Weighted least squares</i> , 220	
8.2.4	<i>Application of constraints</i> , 224	

8.2.5	<i>Equally spaced x-values</i> , 225	
8.3	Piecewise Polynomial Fitting	227
8.3.1	<i>Unsatisfactory fit</i> , 227	
8.3.2	<i>Use of spline functions</i> , 228	
8.3.3	<i>Multiphase polynomial regression</i> , 230	
8.4	Optimal Allocation of Points	231
8.5	Polynomial Regression in Several Variables	234
8.5.1	<i>Surface fitting</i> , 234	
8.5.2	<i>Response surfaces</i> , 234	
	Miscellaneous Exercises 8, 238	
9	ANALYSIS OF VARIANCE	239
9.1	One-Way Classification	239
9.1.1	<i>Representation as a regression model</i> , 239	
9.1.2	<i>Computations</i> , 241	
9.1.3	<i>Expected values</i> , 243	
9.1.4	<i>Reparametrization of the model</i> , 244	
9.1.5	<i>Geometrical considerations</i> , 245	
9.1.6	<i>Identifiability constraints</i> , 247	
9.1.7	<i>Confidence intervals</i> , 247	
9.1.8	<i>Underlying assumptions</i> , 249	
9.1.9	<i>Unequal numbers of observations per mean</i> , 250	
9.2	Two-Way Classification	252
9.2.1	<i>Representation as a regression model</i> , 252	
9.2.2	<i>Test statistics</i> , 254	
9.2.3	<i>Analysis of variance table</i> , 256	
9.2.4	<i>Confidence intervals</i> , 258	
9.2.5	<i>Unequal numbers of observations per mean</i> , 259	
9.3	Higher-Way Classifications with Equal Numbers per Mean	260
9.3.1	<i>Definition of interactions and main effects</i> , 260	
9.3.2	<i>Hypothesis testing</i> , 262	
9.4	Classifications with One Observation per Mean	265
9.4.1	<i>Derivation of test statistics</i> , 265	
9.4.2	<i>Underlying assumptions</i> , 268	
9.4.3	<i>An alternative approach</i> , 271	
9.5	Designs with Simple Block Structure	271
	Miscellaneous Exercises 9, 276	

10	ANALYSIS OF COVARIANCE AND MISSING OBSERVATIONS	279
10.1	Analysis of Covariance	279
10.1.1	<i>Least squares estimation</i> , 279	
10.1.2	<i>Hypothesis testing</i> , 283	
10.1.3	<i>Underlying assumptions</i> , 288	
10.2	Missing Observations	290
10.2.1	<i>Estimation theory</i> , 290	
10.2.2	<i>Hypothesis testing</i> , 293	
10.2.3	<i>Analysis of covariance method</i> , 296	
10.2.4	<i>Missing observations in analysis of covariance models</i> , 300	
	Miscellaneous Exercises 10, 300	
11	COMPUTATIONAL TECHNIQUES FOR FITTING A SPECIFIED REGRESSION	302
11.1	Introduction	302
11.2	Full Rank Case	303
11.2.1	<i>Gaussian elimination</i> , 303	
11.2.2	<i>Cholesky decomposition (square root) method</i> , 304	
11.2.3	<i>Triangular decomposition</i> , 307	
11.2.4	<i>Orthogonal-triangular decomposition</i> , 308	
	a. Gram–Schmidt orthogonalization, 311	
	b. Householder transformations, 312	
	c. Givens transformations, 314	
11.3	Weighted Least Squares	317
11.3.1	<i>The normal equations</i> , 317	
11.3.2	<i>Cholesky method</i> , 317	
11.3.3	<i>Gram–Schmidt method</i> , 318	
11.3.4	<i>Gentleman’s method</i> , 318	
11.4	Comparison of Methods	319
11.5	Less than Full Rank Case	322
11.5.1	<i>Cholesky method for a generalized inverse of $X'X$</i> , 322	
11.5.2	<i>Abbreviated Doolittle method</i> , 324	
11.5.3	<i>Generalized inverse of X using Householder transformations</i> , 324	
11.5.4	<i>Orthogonal decomposition with identifiability constraints</i> , 325	

	11.5.5 <i>Singular value decomposition</i> , 327	
11.6	Improving the Solution by Iterative Refinement	328
11.7	Centering and Scaling the Data	330
	11.7.1 <i>Centering the data</i> , 330	
	11.7.2 <i>Scaling (equilibration)</i> , 334	
11.8	Regression Updating	336
11.9	Adding or Removing a Specified Regressor	338
	11.9.1 <i>Adding a regressor</i> , 338	
	11.9.2 <i>Removing a regressor</i> , 341	
11.10	Hypothesis Testing	342
11.11	Checking Out Computer Programs	344
	Miscellaneous Exercises 11, 346	
12	CHOOSING THE “BEST” REGRESSION	349
12.1	Introduction	349
12.2	Generating All Possible Regressions	349
	12.2.1 <i>Order of generation</i> , 350	
	12.2.2 <i>Method of generation</i> , 351	
	<i>a. Sweeping</i> , 351	
	<i>b. Symmetric sweep</i> , 354	
	<i>c. Furnival’s method</i> , 355	
	<i>d. Householder and Givens transformations</i> , 357	
	12.2.3 <i>Comparing different equations</i> , 358	
	<i>a. Coefficient of determination (R^2)</i> , 359	
	<i>b. Adjusted coefficient of determination</i> , 362	
	<i>c. Mallows’ C_p statistic</i> , 364	
	<i>d. The MSEP criterion</i> , 369	
	<i>e. Other measures</i> , 369	
12.3	Generating the Better Regressions Only	371
	12.3.1 <i>Searching along favorable branches</i> , 371	
	12.3.2 <i>A t-directed search</i> , 374	
12.4	Stepwise Regression	376
	12.4.1 <i>Description of the method</i> , 376	
	12.4.2 <i>Use of sweeping</i> , 377	
	12.4.3 <i>Gauss–Jordan elimination</i> , 379	
	12.4.4 <i>Choice of F values</i> , 380	
	12.4.5 <i>Other one-step methods</i> , 380	
12.5	Other Methods	381

12.5.1	<i>Factor and principal component analysis</i> , 381	
12.5.2	<i>A Bayesian method</i> , 381	
12.6	General Comments	382
	Miscellaneous Exercises 12, 382	
APPENDIX	383
A	Some Matrix Algebra	383
B	Orthogonal Projections	394
C	Normal Probability Paper	398
D	Percentage Points of the Bonferroni t -Statistic	402
E	Distribution of the Largest Absolute Value of k Student t Variates	404
F	Working–Hotelling Confidence Bands for Finite Intervals	411
OUTLINE SOLUTIONS TO EXERCISES	413
REFERENCES	433
INDEX	459

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 $E[Y]$, $\text{var}[Y]$, $\text{cov}[X, Y]$, 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, \dots, d_n and zeros elsewhere is denoted by $\text{diag}(d_1, d_2, \dots, d_n)$, 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, \dots, a_n , we write $\mathbf{a} = [(a_i)]$, and the *length* or *norm* of \mathbf{a} is denoted by $\|\mathbf{a}\|$. Thus

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

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 $\text{tr } \mathbf{A}$ ($= a_{11} + a_{22} + \dots + a_{kk}$, 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