LINEAR MODELS IN STATISTICS

Second Edition

Alvin C. Rencher and G. Bruce Schaalje
Department of Statistics, Brigham Young University, Provo, Utah

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

LINEAR MODELS IN STATISTICS

THE WILEY BICENTENNIAL–KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

William J. Pesce

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

Peter Booth Wiley

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

LINEAR MODELS IN STATISTICS

Second Edition

Alvin C. Rencher and G. Bruce Schaalje
Department of Statistics, Brigham Young University, Provo, Utah

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data:

Rencher, Alvin C., 1934-
Linear models in statistics/Alvin C. Rencher, G. Bruce Schaalje. – 2nd ed.
p. cm.
Includes bibliographical references.
ISBN 978-0-471-75498-5 (cloth)
1. Linear models (Statistics) I. Schaalje, G. Bruce. II. Title.
QA276.R425 2007
519.5′35–dc22

2007024268

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

In the second edition, we have added chapters on Bayesian inference in linear models (Chapter 11) and linear mixed models (Chapter 17), and have upgraded the material in all other chapters. Our continuing objective has been to introduce the theory of linear models in a clear but rigorous format.

In spite of the availability of highly innovative tools in statistics, the main tool of the applied statistician remains the linear model. The linear model involves the simplest and seemingly most restrictive statistical properties: independence, normality, constancy of variance, and linearity. However, the model and the statistical methods associated with it are surprisingly versatile and robust. More importantly, mastery of the linear model is a prerequisite to work with advanced statistical tools because most advanced tools are generalizations of the linear model. The linear model is thus central to the training of any statistician, applied or theoretical.

This book develops the basic theory of linear models for regression, analysis-of-variance, analysis–of–covariance, and linear mixed models. Chapter 18 briefly introduces logistic regression, generalized linear models, and nonlinear models. Applications are illustrated by examples and problems using real data. This combination of theory and applications will prepare the reader to further explore the literature and to more correctly interpret the output from a linear models computer package.

This introductory linear models book is designed primarily for a one-semester course for advanced undergraduates or MS students. It includes more material than can be covered in one semester so as to give an instructor a choice of topics and to serve as a reference book for researchers who wish to gain a better understanding of regression and analysis-of-variance. The book would also serve well as a text for PhD classes in which the instructor is looking for a one-semester introduction, and it would be a good supplementary text or reference for a more advanced PhD class for which the students need to review the basics on their own.

Our overriding objective in the preparation of this book has been *clarity of exposition*. We hope that students, instructors, researchers, and practitioners will find this linear models text more comfortable than most. In the final stages of development, we asked students for written comments as they read each day's assignment. They made many suggestions that led to improvements in readability of the book. We are grateful to readers who have notified us of errors and other suggestions for improvements of the text, and we will continue to be very grateful to readers who take the time to do so for this second edition.

Another objective of the book is to tie up loose ends. There are many approaches to teaching regression, for example. Some books present estimation of regression coefficients for fixed x's only, other books use random x's, some use centered models, and others define estimated regression coefficients in terms of variances and covariances or in terms of correlations. Theory for linear models has been presented using both an algebraic and a geometric approach. Many books present classical (frequentist) inference for linear models, while increasingly the Bayesian approach is presented. We have tried to cover all these approaches carefully and to show how they relate to each other. We have attempted to do something similar for various approaches to analysis-of-variance. We believe that this will make the book useful as a reference as well as a textbook. An instructor can choose the approach he or she prefers, and a student or researcher has access to other methods as well.

The book includes a large number of theoretical problems and a smaller number of applied problems using real datasets. The problems, along with the extensive set of answers in Appendix A, extend the book in two significant ways: (1) the theoretical problems and answers fill in nearly all gaps in derivations and proofs and also extend the coverage of material in the text, and (2) the applied problems and answers become additional examples illustrating the theory. As instructors, we find that having answers available for the students saves a great deal of class time and enables us to cover more material and cover it better. The answers would be especially useful to a reader who is engaging this material outside the formal classroom setting.

The mathematical prerequisites for this book are multivariable calculus and matrix algebra. The review of matrix algebra in Chapter 2 is intended to be sufficiently complete so that the reader with no previous experience can master matrix manipulation up to the level required in this book. Statistical prerequisites include some exposure to statistical theory, with coverage of topics such as distributions of random variables, expected values, moment generating functions, and an introduction to estimation and testing hypotheses. These topics are briefly reviewed as each is introduced. One or two statistical methods courses would also be helpful, with coverage of topics such as t tests, regression, and analysis-of-variance.

We have made considerable effort to maintain consistency of notation throughout the book. We have also attempted to employ standard notation as far as possible and to avoid exotic characters that cannot be readily reproduced on the chalkboard. With a few exceptions, we have refrained from the use of abbreviations and mnemonic devices. We often find these annoying in a book or journal article.

Equations are numbered sequentially throughout each chapter; for example, (3.29) indicates the twenty-ninth numbered equation in Chapter 3. Tables and figures are also numbered sequentially throughout each chapter in the form "Table 7.4" or "Figure 3.2." On the other hand, examples and theorems are numbered sequentially within a section, for example, Theorems 2.2a and 2.2b.

The solution of most of the problems with real datasets requires the use of the computer. We have not discussed command files or output of any particular program, because there are so many good packages available. Computations for the numerical examples and numerical problems were done with SAS. The datasets and SAS command files for all the numerical examples and problems in the text are available on the Internet; see Appendix B.

The references list is not intended to be an exhaustive survey of the literature. We have provided original references for some of the basic results in linear models and have also referred the reader to many up-to-date texts and reference books useful for further reading. When citing references in the text, we have used the standard format involving the year of publication. For journal articles, the year alone suffices, for example, Fisher (1921). But for a specific reference in a book, we have included a page number or section, as in Hocking (1996, p. 216).

Our selection of topics is intended to prepare the reader for a better understanding of applications and for further reading in topics such as mixed models, generalized linear models, and Bayesian models. Following a brief introduction in Chapter 1, Chapter 2 contains a careful review of all aspects of matrix algebra needed to read the book. Chapters 3, 4, and 5 cover properties of random vectors, matrices, and quadratic forms. Chapters 6, 7, and 8 cover simple and multiple linear regression, including estimation and testing hypotheses and consequences of misspecification of the model. Chapter 9 provides diagnostics for model validation and detection of influential observations. Chapter 10 treats multiple regression with random x's. Chapter 11 covers Bayesian multiple linear regression models along with Bayesian inferences based on those models. Chapter 12 covers the basic theory of analysis-of-variance models, including estimability and testability for the overparameterized model, reparameterization, and the imposition of side conditions. Chapters 13 and 14 cover balanced one-way and two-way analysis-of-variance models using an overparameterized model. Chapter 15 covers unbalanced analysis-of-variance models using a cell means model, including a section on dealing with empty cells in two-way analysis-of-variance. Chapter 16 covers analysis of covariance models. Chapter 17 covers the basic theory of linear mixed models, including residual maximum likelihood estimation of variance components, approximate small-sample inferences for fixed effects, best linear unbiased prediction of random effects, and residual analysis. Chapter 18 introduces additional topics such as nonlinear regression, logistic regression, loglinear models, Poisson regression, and generalized linear models.

In our class for first-year master's-level students, we cover most of the material in Chapters 2–5, 7–8, 10–12, and 17. Many other sequences are possible. For example, a thorough one-semester regression and analysis-of-variance course could cover Chapters 1–10, and 12–15.

Al's introduction to linear models came in classes taught by Dale Richards and Rolf Bargmann. He also learned much from the books by Graybill, Scheffé, and Rao. Al expresses thanks to the following for reading the first edition manuscript and making many valuable suggestions: David Turner, John Walker, Joel Reynolds, and Gale Rex Bryce. Al thanks the following students at Brigham Young University (BYU) who helped with computations, graphics, and typing of the first edition: David Fillmore, Candace Baker, Scott Curtis, Douglas Burton, David Dahl, Brenda Price, Eric Hintze, James Liechty, and Joy Willbur. The students

in Al's Linear Models class went through the manuscript carefully and spotted many typographical errors and passages that needed additional clarification.

Bruce's education in linear models came in classes taught by Mel Carter, Del Scott, Doug Martin, Peter Bloomfield, and Francis Giesbrecht, and influential short courses taught by John Nelder and Russ Wolfinger.

We thank Bruce's Linear Models classes of 2006 and 2007 for going through the book and new chapters. They made valuable suggestions for improvement of the text. We thank Paul Martin and James Hattaway for invaluable help with LaTex. The Department of Statistics, Brigham Young University provided financial support and encouragement throughout the project.

Second Edition

For the second edition we added Chapter 11 on Bayesian inference in linear models (including Gibbs sampling) and Chapter 17 on linear mixed models.

We also added a section in Chapter 2 on vector and matrix calculus, adding several new theorems and covering the Lagrange multiplier method. In Chapter 4, we presented a new proof of the conditional distribution of a subvector of a multivariate normal vector. In Chapter 5, we provided proofs of the moment generating function and variance of a quadratic form of a multivariate normal vector. The section on the geometry of least squares was completely rewritten in Chapter 7, and a section on the geometry of least squares in the overparameterized linear model was added to Chapter 12. Chapter 8 was revised to provide more motivation for hypothesis testing and simultaneous inference. A new section was added to Chapter 15 dealing with two-way analysis-of-variance when there are empty cells. This material is not available in any other textbook that we are aware of.

This book would not have been possible without the patience, support, and encouragement of Al's wife LaRue and Bruce's wife Lois. Both have helped and supported us in more ways than they know. This book is dedicated to them.

ALVIN C. RENCHER AND G. BRUCE SCHAALJE

Department of Statistics
Brigham Young University
Provo, Utah

1 Introduction

The scientific method is frequently used as a guided approach to learning. Linear statistical methods are widely used as part of this learning process. In the biological, physical, and social sciences, as well as in business and engineering, linear models are useful in both the planning stages of research and analysis of the resulting data. In Sections 1.1–1.3, we give a brief introduction to simple and multiple linear regression models, and analysis-of-variance (ANOVA) models.

1.1 SIMPLE LINEAR REGRESSION MODEL

In simple linear regression, we attempt to model the relationship between two variables, for example, income and number of years of education, height and weight of people, length and width of envelopes, temperature and output of an industrial process, altitude and boiling point of water, or dose of a drug and response. For a linear relationship, we can use a model of the form

$$y = \beta_0 + \beta_1 x + \varepsilon, \tag{1.1}$$

where y is the *dependent* or *response* variable and x is the *independent* or *predictor* variable. The random variable ε is the error term in the model. In this context, *error* does not mean mistake but is a statistical term representing random fluctuations, measurement errors, or the effect of factors outside of our control.

The linearity of the model in (1.1) is an assumption. We typically add other assumptions about the distribution of the error terms, independence of the observed values of y, and so on. Using observed values of x and y, we estimate β_0 and β_1 and make inferences such as confidence intervals and tests of hypotheses for β_0 and β_1. We may also use the estimated model to forecast or predict the value of y for a particular value of x, in which case a measure of predictive accuracy may also be of interest.

Estimation and inferential procedures for the simple linear regression model are developed and illustrated in Chapter 6.

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

1.2 MULTIPLE LINEAR REGRESSION MODEL

The response y is often influenced by more than one predictor variable. For example, the yield of a crop may depend on the amount of nitrogen, potash, and phosphate fertilizers used. These variables are controlled by the experimenter, but the yield may also depend on uncontrollable variables such as those associated with weather.

A linear model relating the response y to several predictors has the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon. \tag{1.2}$$

The parameters $\beta_0, \beta_1, \ldots, \beta_k$ are called *regression coefficients*. As in (1.1), ε provides for random variation in y not explained by the x variables. This random variation may be due partly to other variables that affect y but are not known or not observed.

The model in (1.2) is linear in the β parameters; it is not necessarily linear in the x variables. Thus models such as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 \sin x_2 + \varepsilon$$

are included in the designation *linear model*.

A model provides a theoretical framework for better understanding of a phenomenon of interest. Thus a model is a mathematical construct that we believe may represent the mechanism that generated the observations at hand. The postulated model may be an idealized oversimplification of the complex real-world situation, but in many such cases, empirical models provide useful approximations of the relationships among variables. These relationships may be either associative or causative.

Regression models such as (1.2) are used for various purposes, including the following:

1. *Prediction*. Estimates of the individual parameters $\beta_0, \beta_1, \ldots, \beta_k$ are of less importance for prediction than the overall influence of the x variables on y. However, good estimates are needed to achieve good prediction performance.
2. *Data Description or Explanation*. The scientist or engineer uses the estimated model to summarize or describe the observed data.
3. *Parameter Estimation*. The values of the estimated parameters may have theoretical implications for a postulated model.
4. *Variable Selection or Screening*. The emphasis is on determining the importance of each predictor variable in modeling the variation in y. The predictors that are associated with an important amount of variation in y are retained; those that contribute little are deleted.
5. *Control of Output*. A cause-and-effect relationship between y and the x variables is assumed. The estimated model might then be used to control the

output of a process by varying the inputs. By systematic experimentation, it may be possible to achieve the optimal output.

There is a fundamental difference between purposes 1 and 5. For prediction, we need only assume that the same correlations that prevailed when the data were collected also continue in place when the predictions are to be made. Showing that there is a significant relationship between y and the x variables in (1.2) does not necessarily prove that the relationship is causal. To establish causality in order to control output, the researcher must choose the values of the x variables in the model and use randomization to avoid the effects of other possible variables unaccounted for. In other words, to ascertain the effect of the x variables on y when the x variables are changed, it is necessary to change them.

Estimation and inferential procedures that contribute to the five purposes listed above are discussed in Chapters 7–11.

1.3 ANALYSIS-OF-VARIANCE MODELS

In analysis-of-variance (ANOVA) models, we are interested in comparing several populations or several conditions in a study. Analysis-of-variance models can be expressed as linear models with restrictions on the x values. Typically the x's are 0s or 1s. For example, suppose that a researcher wishes to compare the mean yield for four types of catalyst in an industrial process. If n observations are to be obtained for each catalyst, one model for the $4n$ observations can be expressed as

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, \ldots, n, \tag{1.3}$$

where μ_i is the mean corresponding to the ith catalyst. A hypothesis of interest is $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. The model in (1.3) can be expressed in the alternative form

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, \ldots, n. \tag{1.4}$$

In this form, α_i is the effect of the ith catalyst, and the hypothesis can be expressed as $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$.

Suppose that the researcher also wishes to compare the effects of three levels of temperature and that n observations are taken at each of the 12 catalyst–temperature combinations. Then the model can be expressed as

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{1.5}$$

$$i = 1, 2, 3, 4; \quad j = 1, 2, 3; \quad k = 1, 2, \ldots, n,$$

where μ_{ij} is the mean for the ijth catalyst–temperature combination, α_i is the effect of the ith catalyst, β_j is the effect of the jth level of temperature, and γ_{ij} is the interaction or joint effect of the ith catalyst and jth level of temperature.

In the examples leading to models (1.3)–(1.5), the researcher chooses the type of catalyst or level of temperature and thus applies different *treatments* to the objects or experimental units under study. In other settings, we compare the means of variables measured on natural groupings of units, for example, males and females or various geographic areas.

Analysis-of-variance models can be treated as a special case of regression models, but it is more convenient to analyze them separately. This is done in Chapters 12–15. Related topics, such as analysis-of-covariance and mixed models, are covered in Chapters 16–17.

2 Matrix Algebra

If we write a linear model such as (1.2) for each of n observations in a dataset, the n resulting models can be expressed in a single compact matrix expression. Then the estimation and testing results can be more easily obtained using matrix theory.

In the present chapter, we review the elements of matrix theory needed in the remainder of the book. Proofs that seem instructive are included or called for in the problems. For other proofs, see Graybill (1969), Searle (1982), Harville (1997), Schott (1997), or any general text on matrix theory. We begin with some basic definitions in Section 2.1.

2.1 MATRIX AND VECTOR NOTATION

2.1.1 Matrices, Vectors, and Scalars

A *matrix* is a rectangular or square array of numbers or variables. We use uppercase boldface letters to represent matrices. In this book, all elements of matrices will be real numbers or variables representing real numbers. For example, the height (in inches) and weight (in pounds) for three students are listed in the following matrix:

$$\mathbf{A} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix}. \tag{2.1}$$

To represent the elements of $\mathbf{A}$ as variables, we use

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}. \tag{2.2}$$

The first subscript in a_{ij} indicates the row; the second identifies the column. The notation $\mathbf{A} = (a_{ij})$ represents a matrix by means of a typical element.

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

The matrix **A** in (2.1) or (2.2) has three rows and two columns, and we say that **A** is 3×2, or that the *size* of **A** is 3×2.

A *vector* is a matrix with a single row or column. Elements in a vector are often identified by a single subscript; for example

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

As a convention, we use lowercase boldface letters for column vectors and lowercase boldface letters followed by the prime symbol (′) for row vectors; for example

$$\mathbf{x}' = (x_1, x_2, x_3) = (x_1\ x_2\ x_3).$$

(Row vectors are regarded as *transposes* of column vectors. The transpose is defined in Section 2.1.3 below). We use either commas or spaces to separate elements of a row vector.

Geometrically, a row or column vector with p elements can be associated with a point in a p-dimensional space. The elements in the vector are the coordinates of the point. Sometimes we are interested in the distance from the origin to the point (vector), the distance between two points (vectors), or the angle between the arrows drawn from the origin to the two points.

In the context of matrices and vectors, a single real number is called a *scalar*. Thus 2.5, -9, and 7.26 are scalars. A variable representing a scalar will be denoted by a lightface letter (usually lowercase), such as c. A scalar is technically distinct from a 1×1 matrix in terms of its uses and properties in matrix algebra. The same notation is often used to represent a scalar and a 1×1 matrix, but the meaning is usually obvious from the context.

2.1.2 Matrix Equality

Two matrices or two vectors are equal if they are of the same size and if the elements in corresponding positions are equal; for example

$$\begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix},$$

but

$$\begin{pmatrix} 5 & 2 & -9 \\ 8 & -4 & 6 \end{pmatrix} \neq \begin{pmatrix} 5 & 3 & -9 \\ 8 & -4 & 6 \end{pmatrix}.$$

2.1.3 Transpose

If we interchange the rows and columns of a matrix $\mathbf{A}$, the resulting matrix is known as the *transpose* of $\mathbf{A}$ and is denoted by $\mathbf{A}'$; for example

$$\mathbf{A} = \begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}, \qquad \mathbf{A}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}.$$

Formally, if $\mathbf{A}$ is denoted by $\mathbf{A} = (a_{ij})$, then $\mathbf{A}'$ is defined as

$$\mathbf{A}' = (a_{ij})' = (a_{ji}). \tag{2.3}$$

This notation indicates that the element in the ith row and jth column of $\mathbf{A}$ is found in the jth row and ith column of $\mathbf{A}'$. If the matrix $\mathbf{A}$ is $n \times p$, then $\mathbf{A}'$ is $p \times n$.

If a matrix is transposed twice, the result is the original matrix.

Theorem 2.1. If $\mathbf{A}$ is any matrix, then

$$(\mathbf{A}')' = \mathbf{A}. \tag{2.4}$$

PROOF. By (2.3), $\mathbf{A}' = (a_{ij})' = (a_{ji})$. Then $(\mathbf{A}')' = (a_{ji})' = (a_{ij}) = \mathbf{A}$. □
(The notation □ is used to indicate the end of a theorem proof, corollary proof or example.)

2.1.4 Matrices of Special Form

If the transpose of a matrix $\mathbf{A}$ is the same as the original matrix, that is, if $\mathbf{A}' = \mathbf{A}$ or equivalently $(a_{ji}) = (a_{ij})$, then the matrix $\mathbf{A}$ is said to be *symmetric*. For example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}$$

is symmetric. Clearly, all symmetric matrices are square.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{22} & \cdots & a_{pp} \end{pmatrix}$$

The *diagonal* of a $p \times p$ square matrix $\mathbf{A} = (a_{ij})$ consists of the elements $a_{11}, a_{22}, \ldots, a_{pp}$. If a matrix contains zeros in all off-diagonal positions, it is said to be a *diagonal matrix*; for example, consider the matrix

$$\mathbf{D} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix},$$

which can also be denoted as

$$\mathbf{D} = \mathrm{diag}(8,\ -3, 0,\ 4).$$

We also use the notation diag($\mathbf{A}$) to indicate a diagonal matrix with the same diagonal elements as $\mathbf{A}$; for example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}, \quad \mathrm{diag}(\mathbf{A}) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

A diagonal matrix with a 1 in each diagonal position is called an *identity* matrix, and is denoted by $\mathbf{I}$; for example

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.5}$$

An *upper triangular matrix* is a square matrix with zeros below the diagonal; for example,

$$\mathbf{T} = \begin{pmatrix} 7 & 2 & 3 & -5 \\ 0 & 0 & -2 & 6 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

A *lower triangular matrix* is defined similarly.

A vector of 1s is denoted by $\mathbf{j}$:

$$\mathbf{j} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{2.6}$$

A square matrix of 1s is denoted by $\mathbf{J}$; for example

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{2.7}$$

We denote a vector of zeros by $\mathbf{0}$ and a matrix of zeros by $\mathbf{O}$; for example

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{O} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.8}$$

2.2 OPERATIONS

We now define sums and products of matrices and vectors and consider some properties of these sums and products.

2.2.1 Sum of Two Matrices or Two Vectors

If two matrices or two vectors are the same size, they are said to be *conformal for addition*. Their *sum* is found by adding corresponding elements. Thus, if $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $n \times p$, then $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is also $n \times p$ and is found as $\mathbf{C} = (c_{ij}) = (a_{ij} + b_{ij})$; for example

$$\begin{pmatrix} 7 & -3 & 4 \\ 2 & 8 & -5 \end{pmatrix} + \begin{pmatrix} 11 & 5 & -6 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 18 & 2 & -2 \\ 5 & 12 & -3 \end{pmatrix}.$$

The *difference* $\mathbf{D} = \mathbf{A} - \mathbf{B}$ between two conformal matrices $\mathbf{A}$ and $\mathbf{B}$ is defined similarly: $\mathbf{D} = (d_{ij}) = (a_{ij} - b_{ij})$.

Two properties of matrix addition are given in the following theorem.

Theorem 2.2a. If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times m$, then

(i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. (2.9)

(ii) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$. (2.10)

□

2.2.2 Product of a Scalar and a Matrix

Any scalar can be multiplied by any matrix. The product of a scalar and a matrix is defined as the product of each element of the matrix and the scalar:

$$c\mathbf{A} = (ca_{ij}) = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}. \tag{2.11}$$

Since $ca_{ij} = a_{ij}c$, the product of a scalar and a matrix is commutative:

$$c\mathbf{A} = \mathbf{A}c. \tag{2.12}$$

2.2.3 Product of Two Matrices or Two Vectors

In order for the product $\mathbf{AB}$ to be defined, the number of columns in $\mathbf{A}$ must equal the number of rows in $\mathbf{B}$, in which case $\mathbf{A}$ and $\mathbf{B}$ are said to be *conformal for multiplication*. Then the (ij)th element of the product $\mathbf{C} = \mathbf{AB}$ is defined as

$$c_{ij} = \sum_k a_{ik}b_{kj}, \tag{2.13}$$

which is the sum of products of the elements in the ith row of $\mathbf{A}$ and the elements in the jth column of $\mathbf{B}$. Thus we multiply every row of $\mathbf{A}$ by every column of $\mathbf{B}$. If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$, then $\mathbf{C} = \mathbf{AB}$ is $n \times p$. We illustrate matrix multiplication in the following example.

Example 2.2.3. Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 6 & 5 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \begin{pmatrix} 2\cdot 1 + 1\cdot 2 + 3\cdot 3 & 2\cdot 4 + 1\cdot 6 + 3\cdot 8 \\ 4\cdot 1 + 6\cdot 2 + 5\cdot 3 & 4\cdot 4 + 6\cdot 6 + 5\cdot 8 \end{pmatrix} = \begin{pmatrix} 13 & 38 \\ 31 & 92 \end{pmatrix},$$

$$\mathbf{BA} = \begin{pmatrix} 18 & 25 & 23 \\ 28 & 38 & 36 \\ 38 & 51 & 49 \end{pmatrix}.$$

□

Note that a 1×1 *matrix* $\mathbf{A}$ can only be multiplied on the right by a $1 \times n$ matrix $\mathbf{B}$ or on the left by an $n \times 1$ matrix $\mathbf{C}$, whereas a *scalar* can be multiplied on the right or left by a matrix of any size.

If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$, where $n \neq p$, then $\mathbf{AB}$ is defined, but $\mathbf{BA}$ is not defined. If $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $p \times n$, then $\mathbf{AB}$ is $n \times n$ and $\mathbf{BA}$ is $p \times p$. In this case, of course, $\mathbf{AB} \neq \mathbf{BA}$, as illustrated in Example 2.2.3. If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$, then $\mathbf{AB}$ and $\mathbf{BA}$ are the same size, but, in general

$$\mathbf{AB} \neq \mathbf{BA}. \tag{2.14}$$

[There are a few exceptions to (2.14), for example, two diagonal matrices or a square matrix and an identity.] Thus matrix multiplication is not commutative, and certain familiar manipulations with real numbers cannot be done with matrices. However, matrix multiplication is distributive over addition or subtraction:

$$\mathbf{A}(\mathbf{B} \pm \mathbf{C}) = \mathbf{AB} \pm \mathbf{AC}, \tag{2.15}$$

$$(\mathbf{A} \pm \mathbf{B})\mathbf{C} = \mathbf{AC} \pm \mathbf{BC}. \tag{2.16}$$

Using (2.15) and (2.16), we can expand products such as $(\mathbf{A} - \mathbf{B})(\mathbf{C} - \mathbf{D})$:

$$\begin{aligned}(\mathbf{A} - \mathbf{B})(\mathbf{C} - \mathbf{D}) &= (\mathbf{A} - \mathbf{B})\mathbf{C} - (\mathbf{A} - \mathbf{B})\mathbf{D} \qquad [\text{by } (2.15)] \\ &= \mathbf{AC} - \mathbf{BC} - \mathbf{AD} + \mathbf{BD} \qquad [\text{by } (2.16)].\end{aligned} \tag{2.17}$$

Multiplication involving vectors follows the same rules as for matrices. Suppose that $\mathbf{A}$ is $n \times p$, $\mathbf{b}$ is $p \times 1$, $\mathbf{c}$ is $p \times 1$, and $\mathbf{d}$ is $n \times 1$. Then $\mathbf{Ab}$ is a column vector of size $n \times 1$, $\mathbf{d}'\mathbf{A}$ is a row vector of size $1 \times p$, $\mathbf{b}'\mathbf{c}$ is a sum of products (1×1), $\mathbf{bc}'$ is a $p \times p$ matrix, and $\mathbf{cd}'$ is a $p \times n$ matrix. Since $\mathbf{b}'\mathbf{c}$ is a 1×1 sum of products, it is equal to $\mathbf{c}'\mathbf{b}$:

$$\begin{aligned}\mathbf{b}'\mathbf{c} &= b_1c_1 + b_2c_2 + \cdots + b_pc_p, \\ \mathbf{c}'\mathbf{b} &= c_1b_1 + c_2b_2 + \cdots + c_pb_p, \\ \mathbf{b}'\mathbf{c} &= \mathbf{c}'\mathbf{b}.\end{aligned} \tag{2.18}$$

The matrix $\mathbf{cd}'$ is given by

$$\mathbf{cd}' = \begin{pmatrix} c_1d_1 & c_1d_2 & \cdots & c_1d_n \\ c_2d_1 & c_2d_2 & \cdots & c_2d_n \\ \vdots & \vdots & & \vdots \\ c_pd_1 & c_pd_2 & \cdots & c_pd_n \end{pmatrix}. \tag{2.19}$$

Similarly

$$\mathbf{b}'\mathbf{b} = b_1^2 + b_2^2 + \cdots + b_p^2, \tag{2.20}$$

$$\mathbf{b}\mathbf{b}' = \begin{pmatrix} b_1^2 & b_1b_2 & \cdots & b_1b_p \\ b_2b_1 & b_2^2 & \cdots & b_2b_p \\ \vdots & \vdots & & \vdots \\ b_pb_1 & b_pb_2 & \cdots & b_p^2 \end{pmatrix}. \tag{2.21}$$

Thus, $\mathbf{b}'\mathbf{b}$ is a sum of squares and $\mathbf{b}\mathbf{b}'$ is a (symmetric) square matrix.

The square root of the sum of squares of the elements of a $p \times 1$ vector $\mathbf{b}$ is the distance from the origin to the point $\mathbf{b}$ and is also referred to as the *length* of $\mathbf{b}$:

$$\text{Length of } \mathbf{b} = \sqrt{\mathbf{b}'\mathbf{b}} = \sqrt{\sum_{i=1}^{p} b_i^2}. \tag{2.22}$$

If $\mathbf{j}$ is an $n \times 1$ vector of 1s as defined in (2.6), then by (2.20) and (2.21), we have

$$\mathbf{j}'\mathbf{j} = n, \quad \mathbf{j}\mathbf{j}' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \mathbf{J}, \tag{2.23}$$

where $\mathbf{J}$ is an $n \times n$ square matrix of 1s as illustrated in (2.7). If $\mathbf{a}$ is $n \times 1$ and $\mathbf{A}$ is $n \times p$, then

$$\mathbf{a}'\mathbf{j} = \mathbf{j}'\mathbf{a} = \sum_{i=1}^{n} a_i, \tag{2.24}$$

$$\mathbf{j}'\mathbf{A} = \left(\sum_i a_{i1}, \sum_i a_{i2}, \ldots, \sum_i a_{ip} \right), \qquad \mathbf{A}\mathbf{j} = \begin{pmatrix} \sum_j a_{1j} \\ \sum_j a_{2j} \\ \vdots \\ \sum_j a_{nj} \end{pmatrix}. \tag{2.25}$$

Thus $\mathbf{a}'\mathbf{j}$ is the sum of the elements in $\mathbf{a}$, $\mathbf{j}'\mathbf{A}$ contains the column sums of $\mathbf{A}$, and $\mathbf{A}\mathbf{j}$ contains the row sums of $\mathbf{A}$. Note that in $\mathbf{a}'\mathbf{j}$, the vector $\mathbf{j}$ is $n \times 1$; in $\mathbf{j}'\mathbf{A}$, the vector $\mathbf{j}$ is $n \times 1$; and in $\mathbf{A}\mathbf{j}$, the vector $\mathbf{j}$ is $p \times 1$.