Wiley Series in Probability and Statistics

Shayle R. Searle and André I. Khuri

Matrix Algebra Useful for Statistics

Second Edition

Matrix Algebra Useful for Statistics

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by *Walter A. Shewhart and Samuel S. Wilks*

The *Wiley Series in Probability and Statistics* is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at http://www.wiley.com/go/wsps

Matrix Algebra Useful for Statistics

Second Edition

Shayle R. Searle
André I. Khuri

WILEY

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

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-1-118-93514-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

In Memory of Shayle R. Searle, a Good Friend and Colleague

To My Faithful Wife, Ronnie, and Dedicated Children, Marcus and Roxanne,
and Their Families

Contents

Preface

The primary objective of the second edition is to update the material in the first edition. This is a significant undertaking given that the first edition appeared in 1982. It should be first pointed out that this is more than just an update. It is in fact a major revision of the material affecting not only its presentation, but also its applicability and use by the reader.

The second edition consists of three parts. Part I is comprised of Chapters 1–9, which with the exception of Chapter 1, covers material based on an update of Chapters 1–12 in the first edition. These chapters are preceded by an introductory chapter giving historical perspectives on matrix algebra. Chapter 1 is new. It discusses vector spaces and linear transformations that represent an introduction to matrices. Part II addresses applications of matrices in statistics. It consists of Chapters 10–14. Chapters 10–11 constitute an update of Chapters 13–14 in the first edition. Chapter 12 is similar to Chapter 15 in the first edition. It covers models that are less than full rank. Chapter 13 is entirely new. It discusses the analysis of balanced linear models using direct products of matrices. Chapter 14 is also a new addition that covers multiresponse linear models where several responses can be of interest. Part III is new. It covers computational aspects of matrices and consists of three chapters. Chapter 15 is on the use of SAS/IML, Chapter 16 covers the use of MATLAB, and Chapter 17 discusses the implementation of R in matrix computations. These three chapters are self-contained and provide the reader with the necessary tools to carry out all the computations described in the book. The reader can choose whichever software he/she feels comfortable with. It is also quite easy to learn new computational techniques that can be beneficial.

The second edition displays a large number of figures to illustrate certain computational details. This provides a visual depiction of matrix entities such as the plotting of a matrix and the graphical representation of a determinant. In addition, many examples have been included to provide a better understanding of the material.

A new feature in the second edition is the addition of detailed solutions to all the odd-numbered exercises. The even-numbered solutions will be placed online by the publisher.

This can be helpful to the reader who desires to use the book as a source for learning matrix algebra.

As with the first edition, the second edition emphasizes the "bringing to a broad spectrum of readers a knowledge of matrix algebra that is useful in the statistical analysis of data and in statistics in general." The second edition should therefore appeal to all those who desire to gain a better understanding of matrix algebra and its applications in linear models and multivariate statistics. The computing capability that the reader needs is particularly enhanced by the inclusion of Part III on matrix computations.

I am grateful to my wife Ronnie, my daughter Roxanne, and son Marcus for their support and keeping up with my progress in writing the book over the past 3 years. I am also grateful to Steve Quigley, a former editor with John Wiley & Sons, for having given me the opportunity to revise the first edition. Furthermore, my gratitude goes to Julie Platt, an Editor-in-Chief with the SAS Institute, for allowing me to use the SAS software in the second edition for two consecutive years.

André I. Khuri

Jacksonville, Florida
January 2017

Preface to the First Edition

Algebra is a mathematical shorthand for language, and matrices are a shorthand for algebra. Consequently, a special value of matrices is that they enable many mathematical operations, especially those arising in statistics and the quantitative sciences, to be expressed concisely and with clarity. The algebra of matrices is, of course, in no way new, but its presentation is often so surrounded by the trappings of mathematical generality that assimilation can be difficult for readers who have only limited ability or training in mathematics. Yet many such people nowadays find a knowledge of matrix algebra necessary for their work, especially where statistics and/or computers are involved. It is to these people that I address this book, and for them, I have attempted to keep the mathematical presentation as informal as possible.

The pursuit of knowledge frequently involves collecting data, and those responsible for the collecting must appreciate the need for analyzing their data to recover and interpret the information contained therein. Such people must therefore understand some of the mathematical tools necessary for this analysis, to an extent either that they can carry out their own analysis, or that they can converse with statisticians and mathematicians whose help will otherwise be needed. One of the necessary tools is matrix algebra. It is becoming as necessary to science today as elementary calculus has been for generations. Matrices originated in mathematics more than a century ago, but their broad adaptation to science is relatively recent, prompted by the widespread acceptance of statistical analysis of data, and of computers to do that analysis; both statistics and computing rely heavily on matrix algebra. The purpose of this book is therefore that of bringing to a broad spectrum of readers a knowledge of matrix algebra that is useful in the statistical analysis of data and in statistics generally.

The basic prerequisite for using the book is high school algebra. Differential calculus is used on only a few pages, which can easily be omitted; nothing will be lost insofar as a general understanding of matrix algebra is concerned. Proofs and demonstrations of most of the theory are given, for without them the presentation would be lifeless. But in every chapter the theoretical development is profusely illustrated with elementary numerical examples and with illustrations taken from a variety of applied sciences. And the last three chapters are devoted solely to uses of matrix algebra in statistics, with Chapters 14 and 15 outlining two of the most widely used statistical techniques: regression and linear models.

The mainstream of the book is its first 11 chapters, beginning with one on introductory concepts that includes a discussion of subscript and summation notation. This is followed by four chapters dealing with basic arithmetic, special matrices, determinants and inverses. Chapters 6 and 7 are on rank and canonical forms, 8 and 9 deal with generalized inverses and solving linear equations, 10 is a collection of results on partitioned matrices, and 11 describes eigenvalues and eigenvectors. Background theory for Chapter 11 is collected in an appendix, Chapter 11A, some summaries and miscellaneous topics make up Chapter 12, statistical illustrations constitute Chapter 13, and Chapters 14 and 15 describe regression and linear models. All chapters except the last two end with exercises.

Occasional sections and paragraphs can be omitted at a first reading, especially by those whose experience in mathematics is somewhat limited. These portions of the book are printed in small type and, generally speaking, contain material subsidiary to the main flow of the text—material that may be a little more advanced in mathematical presentation than the general level otherwise maintained.

Chapters, and sections within chapters, are numbered with Arabic numerals 1, 2, 3,… Within-chapter references to sections are by section number, but references across chapters use the decimal system, for example, Section 1.3 is Section 3 of Chapter 1. These numbers are also shown in the running head of each page, for example, [1.3] is found on page 4. Numbered equations are (1), (2),…, within each chapter. Those of one chapter are seldom referred to in another, but when they are, the chapter reference is explicit; otherwise "equation (3)" or more simply "(3)" means the equation numbered (3) in the chapter concerned. Exercises are in unnumbered sections and are referenced by their chapter number; for example, Exercise 6.2 is Exercise 2 at the end of Chapter 6.

I am greatly indebted to George P. H. Styan for his exquisitely thorough readings of two drafts of the manuscript and his extensive and very helpful array of comments. Harold V. Henderson's numerous suggestions for the final manuscript were equally as helpful. Readers of *Matrix Algebra for the Biological Sciences* (Wiley, 1966), and students in 15 years of my matrix algebra course at Cornell have also contributed many useful ideas. Particular thanks go to Mrs. Helen Seamon for her superb accuracy on the typewriter, patience, and fantastic attention to detail; such attributes are greatly appreciated.

SHAYLE R. SEARLE

Ithaca, New York
May 1982

Introduction

Historical Perspectives on Matrix Algebra

It is difficult to determine the origin of matrices from the historical point of view. Given the association between matrices and simultaneous linear equations, it can be argued that the history of matrices goes back to at least the third century BC. The Babylonians used simultaneous linear equations to study problems that pertained to agriculture in the fertile region between the Tigris and Euphrates rivers in ancient Mesopotamia (present day Iraq). They inscribed their findings, using a wedge-shaped script, on soft clay tablets which were later baked in ovens resulting in what is known as cuneiform tablets (see Figures 1 and 2).

This form of writing goes back to about 3000 BC (see Knuth, 1972). For example, a tablet dating from around 300 BC was found to contain a description of a problem that could be formulated in terms of two simultaneous linear equations in two variables. The description referred to two fields whose total area, the rate of production of grain per field, and their total yield were all given. It was required to determine the area of each field (see O'Connor and Robertson, 1996). The ancient Chinese also dealt with simultaneous linear equations between 200 BC and 100 BC in studying, for example, corn production. In fact, the text, *Nine Chapters on the Mathematical Art*, which was written during the Han Dynasty, played an important role in the development of mathematics in China. It was a practical handbook of mathematics consisting of 246 problems that pertained to engineering, surveying, trade, and taxation issues (see O'Connor and Robertson, 2003).

The modern development of matrices and matrix algebra did not materialize until the nineteenth century with the work of several mathematicians, including Augustin-Louis Cauchy, Ferdinand Georg Frobenius, Carl Friedrich Gauss, Arthur Cayley, and James Joseph Sylvester, among others. The use of the word "matrix" was first introduced by Sylvester in 1850. This terminology became more common after the publication of Cayley's (1858) memoir on the theory of matrices. In 1829, Cauchy gave the first valid proof that the eigenvalues of a symmetric matrix must be real. He was also instrumental in creating the theory of determinants in his 1812 memoir. Frobenius (1877) wrote an important monograph in which he provided a unifying theory of matrices that combined the work of

Figure 0.1 A Cuneiform Tablet with 97 Linear Equations (YBC4695-1). Yale Babylonian Collection, Yale University Library, New Haven, CT.

several other mathematicians. Hawkins (1974) described Frobenius' paper as representing "an important landmark in the history of the theory of matrices." Hawkins (1975) discussed Cauchy's work and its historical significance to the consideration of algebraic eigenvalue problems during the 18th century.

Science historians and mathematicians have regarded Cayley as the founder of the theory of matrices. His 1858 memoir was considered "the foundation upon which other mathematicians were able to erect the edifice we now call the theory of matrices" (see Hawkins, 1974, p. 561). Cayley was interested in devising a contracted notation to represent a system of m linear equations in n variables of the form

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \quad i = 1, 2, \ldots, m,$$

where the a_{ij}'s are given as coefficients. Cayley and other contemporary algebraists proposed replacing the m equations with a single matrix equation such as

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_m \end{bmatrix}.$$

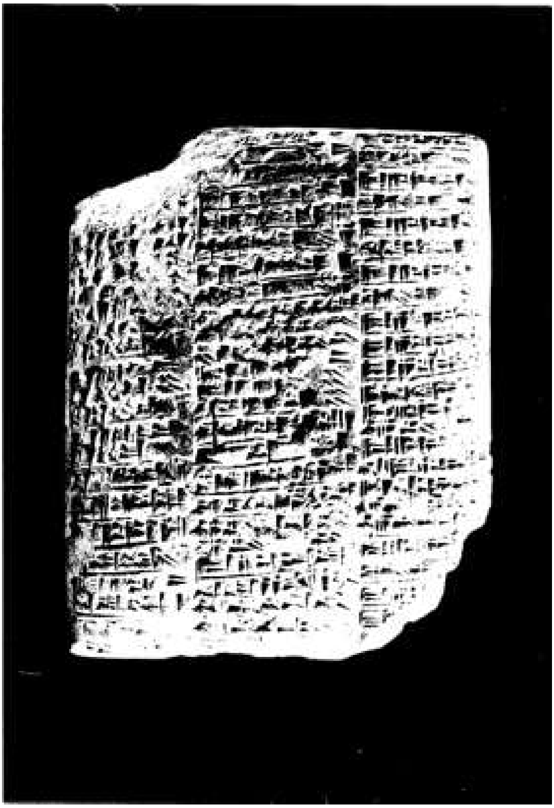


Figure 0.2 *An Old Babylonian Mathematical Text with Linear Equations (YBC4695-2). Yale Babylonian Collection, Yale University Library, New Haven, CT.*

Cayley regarded such a scheme as an operator acting upon the variables, $x_1, x_2, \ldots, x_n$ to produce the variables $y_1, y_2, \ldots, y_m$. This is a multivariable extension of the action of the single coefficient a upon x to produce ax, except that the rules associated with such an extension are different from the single variable case. This led to the development of the algebra of matrices.

Even though Cayley left his mark on the history of matrices, it should be pointed that his role in this endeavor was perhaps overrated by historians to the point of eclipsing the contribution of other mathematicians in the eighteenth and nineteenth centuries. Hawkins (1974) indicated that the ideas Cayley expressed in his 1858 memoir were not particularly original. He cited work by Laguerre (Edmond Nicolas Laguerre), Frobenius, and other mathematicians who had developed similar ideas during the same period, but without a knowledge of Cayley's memoir. This conclusion was endorsed by Farebrother (1997) and Grattan-Guiness (1994). It is perhaps more accurate to conclude, as Hawkins (1975, p. 570) did, that "the history of matrix theory involved the efforts of many mathematicians, that it was indeed an international undertaking." Higham (2008) provided an interesting commentary on the work of Cayley and Sylvester. He indicated that the multi-volume collected works of Cayley and Sylvester were both freely available online at the University of Michigan Historical Mathematics Collection by using the URL, http://quod.lib.umich.edu/u/umhistmath/

(for Cayley, use http://name.umdl.umich.edu/ABS3153.0013.001, and for Sylvester, use http://name.umdl.umich.edu/AAS8085.0002.001).

The history of determinants can be traced to methods used by the ancient Chinese and Japanese to solve a system of linear equations. Seki Kōwa, a distinguished Japanese mathematician of the seventeenth century, discovered the expansion of a determinant in solving simultaneous equations (see, e.g., Smith, 1958, p. 440). However, the methods used by the Chinese and the Japanese did not resemble the methods used nowadays in dealing with determinants. In the West, the theory of determinants is believed to have originated with the German mathematician, Gottfried Leibniz, in the seventeenth century, several years after the work of Seki Kōwa. However, the actual development of this theory did not begin until 1750 with the publication of the book by Gabriel Cramer. In fact, the method of solving a system of n linear equations in n unknowns by means of determinants is known as Cramer's rule. The term "determinant" was first introduced by Gauss in 1801 in connection with quadratic forms. In 1812, Cauchy developed the theory of determinants as is known today. Cayley was the first to introduce the present-day notation of a determinant, namely, of vertical bars enclosing a square matrix, in a paper he wrote in 1841. So, just as in the case of matrices, the history of determinants was an international undertaking shaped by the efforts of many mathematicians. For more interesting facts about the history of determinants, see Miller (1930) and Price (1947).

THE INTRODUCTION OF MATRICES INTO STATISTICS

The entry of matrices into statistics was slow. Farebrother (1999) indicated that matrix algebra was not to emerge until the early part of the twentieth century. Even then, determinants were used in place of matrices in solving equations which were written in longhand. Searle (2000, p. 25) indicated that the year 1930 was a good starting point for the entry of matrices into statistics. That was the year of Volume 1 of the *Annals of Mathematical Statistics*, its very first paper, Wicksell (1930), being "Remarks on Regression." The paper considered finding the least-squares estimates for a linear regression model with one independent variable. The normal equations for getting the model's parameter estimates were expressed in terms of determinants only. No matrices were used. Today, such a subject would have been replete with matrices. Lengthy arguments and numerous equations were given to describe computational methods for general regression models, even in some of the books that appeared in the early 1950s. The slowness of the use of matrices in statistics was partially attributed to the difficulty in producing numerical results in situations involving, for example, regression models with several variables. In particular, the use of a matrix inverse posed a considerable computational difficulty before the advent of computers which came about in only the last 50 years. Today, such computational tasks are carried out quickly and effortlessly for a matrix of a reasonable size using a computer software. During his graduate student days at Cornell University in 1959, Searle (2000) recalled the great excitement he and other classmates in a small computer group had felt when they inverted a 10-by-10 matrix in 7 minutes. At that time this was considered a remarkable feat considering that only a year or two earlier, a friend had inverted a 40-by-40 matrix by hand using electric Marchant or Monroe calculators. That task took 6 weeks! An early beginning to more advanced techniques to inverting a matrix was the Doolittle method, as was described in Anderson and Bancroft (1952, Chapter 15). It is interesting to note that this method was introduced in the U.S. Coast and Geodetic Survey Report of 1878 (see Doolittle, 1881).

Alexander Craig Aitken made important contributions to promoting the use of matrix algebra in statistics in the 1930s. He was a brilliant mathematician from New Zealand with a phenomenal mental capability. It was reported that he could recite the irrational number π to 707 decimal places and multiply two nine-digit numbers in his head in 30 seconds. He was therefore referred to as the Human Calculator. His research, which dealt with matrices and statistics, gave a strong impetus for using matrix algebra as a tool in the development of statistics. This was demonstrated in his work on the optimality properties of the generalized least-squares estimator as in Aitken (1935) (see Farebrother, 1997). The book by Turnbull and Aitken (1932) contained several applications of matrices to statistics, including giving the normal equations for a linear regression model presented in a format still used today. Aitken also did some pioneering work that provided the basis for several important theorems in linear models as in Aitken (1940) with regard to the independence of linear and quadratic forms, and Aitken (1950) concerning the independence of two quadratic forms, both under the assumption of normally distributed variables. The following passage given in Ledermann (1968, pp. 164–165) and referred to in Farebrother (1997, p. 6) best describes Aitken's drive to popularize matrix algebra:

> *With his flair for elegant formalism, Aitken was quick to realize the usefulness of matrix algebra as a powerful tool in many branches of mathematics. At a time when matrix techniques were not yet widely known he applied matrix algebra with striking success to certain statistical problems.*

The use of matrix algebra in statistics began to take hold in the 1940s with the publication of important books by prominent statisticians as in Kendall (1946) with his discussion of the multivariate analysis, and Cramér (1946) who devoted a chapter on matrices, determinants, and quadratic forms. More statistics books and articles utilizing matrix algebra appeared in the 1950s such as Kempthorne (1952) who presented proofs of certain results in linear models using the matrix approach, and the classic book by Anderson (1958), which was one of the early books to heavily depend on matrices in its treatment of multivariate analysis. There were also several statistics papers where matrix algebra played an important role in the presentation. Among these were a number of papers dealing with the independence of quadratic forms, independence of linear and quadratic forms, and the distribution of a quadratic form under the assumption of normally distributed random variables as in Cochran (1934), Craig (1938, 1943), Aitken (1940, 1950), and Ogawa (1949, 1950). The necessary conditions for the independence of quadratic forms and for a quadratic form to have the chi-squared distribution are topics of great interest that are still being pursued in the more recent literature; see, for example, Driscoll and Krasnicka (1995), Olkin (1997), Driscoll (1999), and Khuri (1999). Bartlett (1947) made considerable use of matrices and vector algebra in his paper on multivariate analysis. Dwyer and MacPhail (1948) concentrated on discussing matrix derivatives. The problem of inverting a matrix with minimum computation and high accuracy was of special interest. For example, Hotelling (1943) gave an overview of this area, Goldstine and von Neumann (1951) covered numerical inversion of matrices of high order, and Waugh and Dwyer (1945) discussed compact and efficient methods for the computation of the inverse of a matrix. The distributions of the eigenvalues of a random matrix was also of interest. Mood (1951) outlined a method for deriving the distributions of the eigenvalues of a second-order moment matrix based on a sample from a multivariate normal distribution. Fisher (1939) and Hsu (1939) presented an expression for the density function representing the joint distribution of the eigenvalues of a Wishart distribution. By definition, If $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ are independently and identically distributed as a multivariate normal with mean $\mathbf{0}$ and a variance-covariance matrix $\boldsymbol{\Sigma}$, then the random matrix $\boldsymbol{W}$ defined as

$$\boldsymbol{W} = \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i'$$

is said to have the Wishart distribution with m degrees of freedom. This is written symbolically as $\boldsymbol{W} \sim W_d(\mathbf{0}, \boldsymbol{\Sigma})$, where d is the number of elements in $\boldsymbol{x}_i$, $i = 1, 2, \ldots, m$. The well-known largest root test statistic, λ_{max}, which plays an important role in multivariate analysis of variance, is the maximum eigenvalue of $\boldsymbol{HE}^{-1}$, where $\boldsymbol{H}$ and $\boldsymbol{E}$ are matrices with independent Wishart distributions. The distribution function of λ_{max} was investigated by Roy (1939). Other related test statistics include Wilks' likelihood ratio, $\frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|}$ introduced by Wilks (1932), where $|.|$ denotes the determinant of a matrix, Lawley–Hotelling's trace, $tr(\boldsymbol{HE}^{-1})$ introduced by Lawley (1938) and whose distribution was obtained by Hotelling (1951) (the trace of a square matrix is the sum of its diagonal elements), and Pillai's trace, $tr[\boldsymbol{H}(\boldsymbol{E}+\boldsymbol{H})^{-1}]$ proposed by Pillai (1955). These multivariate test statistics are covered in detail in, for example, Seber (1984).

The singular-value decomposition (SVD) and the generalized inverse (GI) of a matrix (to be defined later in Sections 7.4 and 8.2, respectively) are two important areas in regression and analysis of variance. They were, however, slow to be incorporated into the statistical literature. Rao (1962) used the term "pseudo inverse" to refer to a GI of a rectangular matrix or a square matrix whose determinant is zero (singular matrix). He indicated that such an inverse provided a solution to the normal equations of a less-than-full-rank linear model. This process made it possible to have a unified approach to least-squares estimation, including the case when the matrix of the normal equations was singular. Rao's work was very helpful in this regard. The term pseudoinverse was also used in Greville (1957). Good (1969) emphasized the usefulness of the SVD and described several of its applications in statistics. He indicated that its significant role had been much underrated since it was mentioned in very few statistics books. Mandel (1982) presented a discussion of the use of the SVD in multiple linear regression. Eubank and Webster (1985) pointed out a certain characterization of the GI that could be derived from the SVD. They used this characterization to explain many properties of least-squares estimation. The SVD was also applied to least squares principal component analysis by Whittle (1952). Both the SVD and GI provide helpful tools to the area of least-squares estimation, particulary in situations where the fitted model is of less than full rank, or when the values of the predictor variables of the fitted model fall within a narrow range resulting in the so-called ill-conditioning or multicollinearity in the columns of the matrix, $\boldsymbol{X}$, associated with the model. A generalized inverse can yield a solution to the normal equations of a less-than-full-rank model which can be used to obtain a unique estimate of an estimable linear function of the model's parameters (to be defined in Section 12.10). Statisticians have long recognized that the presence of multicollinearity in the fitted linear model can cause the least-squares estimates of the model's parameters to become unstable. This means that small changes in the response data can result in least-squares estimates with large variances, which is undesirable. To reduce this instability, Hoerl (1962) proposed adding a positive quantity, k, to the diagonal elements of the matrix $\boldsymbol{X}'\boldsymbol{X}$ and using the resulting matrix in place of $\boldsymbol{X}'\boldsymbol{X}$ in the normal equations, where $\boldsymbol{X}'$ is the transpose of $\boldsymbol{X}$. This process led to what became known as ridge regression which was discussed in detail in Hoerl and Kennard (1970). Its introduction generated a large number of articles on this subject in the 1970s and 1980s (see, e.g.,, Smith and Campbell, 1980). Piegorsch and Casella (1989) discussed the early motivation for and development of diagonal increments to lessen the effect of ill conditioning on least-squares estimation. They indicated that one of the basic problems that had led to the use of matrix diagonal increments was "the improvement of a nonlinear least-squares solution when the usual methods fail to provide acceptable estimates." This was first discussed by Levenberg (1944) and later by Marquardt (1963).

An expository article describing some details about the history of the "infusion of matrices into statistics" was given by Searle (2000). See also David's (2006) brief note on a related history. Lowerre (1982) provided an introduction to the SVD and the Moore–Penrose inverse (a particular type of GI to be defined in Section 8.1). Searle (1956) showed how to apply matrix algebra to deriving the sampling variances of the least-squares estimates and the large-sample variances of the maximum likelihood estimates of the variance components in an unbalanced one-way classification model with random effects. This was one of the early papers on using matrices in the analysis of variance components. In "A Conversation with Shayle R. Searle" in Wells' (2009), the author recounted his interview with Searle in which the latter reminisced about his early days as a statistician and his advocacy of using matrices in statistics. See also Searle's interesting "Comments from Thirty Years of Teaching Matrix Algebra to Applied Statisticians" in Searle (1999).

It is important to point out here that an International Workshop on Matrices and Statistics (IWMS) has been organized on an almost annual basis since 1990 (with the exception of 1991 and 1993) to foster the interaction of researchers in the interface between statistics and matrix theory. Puntanen and Styan (2011) gave a short history of the IWMS from its inception in 1990 through 2013 (except for 1991 and 1993). In addition, the authors established an open-access website for the IWMS at the University of Tampere, Finland, which is available at http://www.sis.uta.fi/tilasto/iwms. The PDF file of the authors' article can be downloaded from http://www.sis.uta.fi/tilasto/iwms/IWMS-history.pdf. The twenty-third IWMS was scheduled to be held in Ljubljana, Slovenia, in June 2014.

There are several linear algebra journals with interesting matrix results that can be used in statistics as in *Linear Algebra and Its Applications*, *Linear and Multilinear Algebra*, *Numerical Linear Algebra with Applications*, in addition to *Mathematics Magazine*. The first journal has special issues on linear algebra and statistics. Many papers in these issues were presented at previous meetings of the IWMS. An extensive bibliography on matrices and inequalities with statistical applications was given by Styan and Puntanen (1993).

REFERENCES

Aitken, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinb.*, 55, 42–47.

Aitken, A. C. (1940). On the independence of linear and quadratic forms in samples of normally distributed variates. *Proc. Roy. Soc. Edinb.*, 60, 40–46.

Aitken, A. C. (1950). On the statistical independence of quadratic forms in normal variates. *Biometrika*, 37, 93–96.

Anderson, R. L. and Bancroft, T. A. (1952). *Statistical Theory in Research*. McGraw-Hill, New York.

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

Bartlett, M. S. (1947). Multivariate analysis (with discussion). *J. Roy. Stat. Soc. Ser. B*, 9, 176–197.

Cayley, A. (1858). Memoir on the theory of matrices. *Phil. Trans. Roy. Soc. London*, 148, 17–37.

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Camb. Phil. Soc.*, 30, 178–191.

Craig, A. T. (1938). On the independence of certain estimates of variance. *Ann. Math. Stat.*, 9, 48–55.

Craig, A. T. (1943). Note on the independence of certain quadratic forms. *Ann. Math. Stat.*, 14, 195–197.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ.

David, H. A. (2006). The introduction of matrix algebra into statistics. *Am. Stat.*, 60, 162.

Doolittle, M. H. (1881). Method employed in the solution of normal equations and the adjustment of a triangulation. In: Report of the Superintendent of the U. S. Coast Guard and Geodetic Survey Showing the Progress of the Work During the Fiscal Year Ending With June 1878, pp. 115–120. Government Printing Office, Washington, D. C.

Driscoll, M. F. (1999). An improved result relating quadratic forms and chi-square distributions. *Am. Stat.*, 53, 273–275.

Driscoll, M. F. and Krasnicka, B. (1995). An accessible proof of Craig's theorem in the general case. *Am. Stat.*, 49, 59–62.

Dwyer, P. S. and MacPhail, M. S. (1948). Symbolic matrix derivatives. *Ann. Math. Stat.*, 19, 517–534.

Eubank, R. L. and Webster, J. T. (1985). The singular-value decomposition as a tool for solving estimability problems. *Am. Stat.*, 39, 64–66.

Farebrother, R. W. (1997). A. C. Aitken and the consolidation of matrix theory. *Linear Algebra Appl.*, 264, 3–12.

Farebrother, R. W. (1999). *Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900*. Springer-Verlag, New York.

Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen.*, 9, 238–249.

Frobenius, G. (1877). Ueber das Pfaffsche Problem. *J. Reine Angew. Math.*, 82, 230–315.

Goldstine, H. H. and von Neumann, J. (1951). Numerical inverting of matrices of high order. *Proc. Am. Math. Soc.*, 2, 188–202.

Good, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics*, 11, 823–831; Errata: 1970, vol. 12, p. 722.

Grattan-Guiness, I. (1994). A new type of question. On the prehistory of linear and nonlinear programming. *Hist. Modern Math.*, 3, 43–89.

Greville, T. N. (1957). The pseudo inverse of a rectangular or singular matrix and its application to the solution of systems of linear equations. *SIAM Newsletter*, 5, 3–6.

Hawkins, T. W. (1974). The theory of matrices in the 19th century. In: Proceedings of the International Congress of Mathematicians, Vancouver, Canada, vol. 2, pp. 561–570.

Hawkins, T. W. (1975). Cauchy and the spectral theory of matrices. *Historia Mathematica*, 2, 1–29.

Higham, N. J. (2008). Cayley, Sylvester, and early matrix theory. *Linear Algebra Appl.*, 428, 39–43.

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chem. Eng. Prog.*, 60, 54–59.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12, 55–67.

Hotelling, H. (1943). Some new methods in matrix calculation. *Ann. Math. Stat.*, 14, 1–43.

Hotelling, H. (1951). A generalised T-test and measure of multivariate dispersion. *Proc. Second Berkeley Symp. Math. Stat. Prob.*, 1, 23–41.

Hsu, P. L. (1939). On the distribution of the roots of certain determinantal equations. *Ann. Eugen.*, 9, 250–258.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. John Wiley & Sons, New York.

Kendall, M. G. (1946). *The Advanced Theory of Statistics*, Vol. II. Griffin, London.

Khuri, A. I. (1999). A necessary condition for a quadratic form to have a chi-squared distribution: An accessible proof. *Int. J. Math. Educ. Sci. Technol.*, 30, 335–339.

Knuth, D.E. (1972). Ancient Babylonian algorithms. *Commun. Assoc. Comput. Mach.*, 15, 671–677.

Lawley, D. N. (1938). A generalisation of Fisher's Z-test. *Biometrika*, 30, 180–187.

Ledermann, W. W. (1968). A. C. Aitken's work in pure mathematics. *Proc. Edinb. Math. Soc. Ser. 2*, 16, 162–165.

Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quart. Appl. Math.*, 2, 164–168.

Lowerre, J. M. (1982). An introduction to modern matrix methods and statistics. *Am. Stat.*, 36, 113–115.

Mandel, J. (1982). Use of the singular value decomposition in regression analysis. *Am. Stat.*, 36, 15–24.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, 11, 431–441.

Miller, G. A. (1930). On the history of determinants. *Am. Math. Month.*, 37, 216–219.

Mood, A. M. (1951). On the distribution of the characteristic roots of normal second-moment matrices. *Ann. Math. Stat.*, 22, 266–273.

O'Connor, J. J. and Robertson, E. F. (1996). Matrices and determinants. In: http:// www-history.mcs.st-and.ac.uk/HistTopics/Matrices _ and _ determinants.html, the MacTutor History of Mathematics Archive, a website maintained by the School of Mathematics and Statistics, University of St. Andrews, Scotland, UK.

O'Connor, J. J. and Robertson, E.F. (2003). Nine chapters on the mathematical art. In: http://www-history.mcs.st-and.ac.uk/HistTopics/Nine_chapters.html, the MacTutor History of Mathematics Archive, a website maintained by the School of Mathematics and Statistics, University of St. Andrews, Scotland, UK.

Ogawa, J. (1949). On the independence of bilinear and quadratic forms of a random sample from a normal population. *Ann. Inst. Stat. Math.*, 1, 83–108.

Ogawa, J. (1950). On the independence of quadratic forms in a non-central normal system. *Osaka Math. J.*, 2, 151–159.

Olkin, I. (1997). A determinantal proof of the Craig-Sakamoto theorem. *Linear Algebra Appl.*, 264, 217–223.

Piegorsch, W. W. and Casella, G. (1989). The early use of matrix diagonal elements in statistical problems. *SIAM Rev.*, 31, 428–434.

Pillai, K. C. S. (1955). Some new test criterial in multivariate analysis. *Ann. Math. Stat.*, 26, 117–121.

Price, G. B. (1947). Some identities in the theory of determinants. *Am. Math. Month.*, 54, 75–90.

Puntanen, S. and Styan, G. P. H. (2011). A short history of the International Workshop on Matrices and Statistics (IWMS). In: Abstract Book of the 20th IWMS and the 9th Tartu Conference on Multivariate Statistics, Tartu, Estonia, pp. 87–97, edited by T. Kollo, K. Sander, and A. Kaasik.

Rao, C. R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Stat. Soc. Ser. B*, 24, 152–158.

Roy, S. N. (1939). p-Statistics, or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhya*, 4, 381–396.

Searle, S. R. (1956). Matrix methods in components of variance and covariance analysis. *Ann. Math. Stat.*, 27, 737–748.

Searle, S. R. (1999). Comments from thirty years of teaching matrix algebra to applied statisticians. *Image: Bull. Int. Linear Algebra Soc.*, 22, 3–5.

Searle, S. R. (2000). The infusion of matrices into statistics. *Image: Bull. Int. Linear Algebra Soc.*, 24, 25–32.

Seber, G. A. F. (1984). *Multivariate Observations*. John Wiley & Sons, New York.

Smith, D. E. (1958). *History of Mathematics*, Vol. I. Dover, New York.

Smith, G. and Campbell, F. (1980). A critique of some ridge regression estimators. *J. Am. Stat. Assoc.*, 75, 74–81.

Styan, G. P. H. and Puntanen, S. (1993). A guide to books on matrices and on inequalities, with statistical and other applications. Report A283, Department of Mathematical Sciences, University of Tampere, Tampere, Finland.

Turnbull, H. W. and Aitken, A. C. (1932). *An Introduction to the Theory of Canonical Matrices*. Blackie & Sons, London, UK.

Waugh, F.V. and Dwyer, P. S. (1945). Compact computation of the inverse of a matrix. *Ann. Math. Stat.*, 16, 259–271.

Wells, M. T. (2009). A conversation with Shayle R. Searle. *Stat. Sci.*, 24, 244–254.

Whittle, P. (1952). On principal components and least squares methods of factor analysis. *Scand. Actuarial J.*, 35, 223–239.

Wicksell, S. D. (1930). Remarks on regression. *Ann. Math. Stat.*, 1, 3–13.

Wilks, S. S. (1932). Certain generalisations in the analysis of variance. *Biometrika*, 24, 471–494.

About the Companion Website

This book is accompanied by a companion website:

www.wiley.com/go/searle/matrixalgebra2e

The website includes:

- Solutions to even numbered exercises (for instructors only)

Part I

Definitions, Basic Concepts, and Matrix Operations

This is the first of three parts that make up this book. The purpose of Part I is to familiarize the reader with the basic concepts and results of matrix algebra. It is designed to provide the tools needed for the understanding of a wide variety of topics in statistics where matrices are used, such as linear models and multivariate analysis, among others. Some of these topics will be addressed in Part II. Proofs of several theorems are given as we believe that understanding the development of a proof can in itself contribute to acquiring a greater ability in dealing with certain matrix intricacies that may be encountered in statistics. However, we shall not attempt to turn this part into a matrix theory treatise overladen with theorems and proofs, which can be quite insipid. Instead, emphasis will be placed on providing an appreciation of the theory, but without losing track of the objective of learning matrix algebra, namely acquiring the ability to apply matrix results in statistics. The theoretical development in every chapter is illustrated with numerous examples to motivate the learning of the theory. The material in Part II will demonstrate the effectiveness of using such theory in statistics.

Part I consists of the following nine chapters:

Chapter 1: Vector Spaces, Subspaces, and Linear Transformations.
Matrix algebra had its foundation in simultaneous linear equations which represented a linear transformation from one n-dimensional Euclidean space to another of the same dimension. This idea was later extended to include linear transformations between more general spaces, not necessarily of the same dimension. Such linear transformations gave rise to matrices. An n-dimensional Euclidean space is a special case of a wider concept called a vector space.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

Chapter 2: Matrix Notation and Terminology.
In order to understand and work with matrices, it is necessary to be quite familiar with the notation and system of terms used in matrix algebra. This chapter defines matrices as rectangular or square arrays of numbers arranged in rows and columns.

Chapter 3: Determinants.
This chapter introduces determinants and provides a description of their basic properties. Various methods of determinantal expansions are included.

Chapter 4: Matrix Operations.
This chapter covers various aspects of matrix operations such as partitioning of matrices, multiplication, direct sum, and direct products of matrices, the inverse and rank of matrices, and full-rank factorization.

Chapter 5: Special Matrices.
Certain types of matrices are frequently used in statistics, such as symmetric, orthogonal, idempotent, positive definite matrices. This chapter also includes different methods to parameterize orthogonal matrices.

Chapter 6: Eigenvalues and Eigenvectors.
A detailed study is given of the eigenvalues and eigenvectors of square matrices, their properties and actual computation. Eigenvalues of certain special matrices, such as symmetric, orthogonal, and idempotent matrices, are discussed, in addition to those that pertain to direct products and direct sums of matrices.

Chapter 7: Diagonalization of Matrices.
Different methods are given to diagonalize matrices that satisfy certain properties. The Cayley–Hamilton theorem, and the singular-value decomposition of matrices are also covered.

Chapter 8: Generalized Inverses.
The Moore–Penrose inverse and the more general generalized inverses of matrices are discussed. Properties of generalized inverses of symmetric matrices are studied, including the special case of the $\boldsymbol{X}'\boldsymbol{X}$ matrix.

Chapter 9: Matrix Calculus.
Coverage is given of calculus results associated with matrices, such as functions of matrices, infinite series of matrices, vectors of differential operators, quadratic forms, differentiation of matrices, traces, and determinants, in addition to matrices of second-order partial derivatives, and matrix inequalities.

1

Vector Spaces, Subspaces, and Linear Transformations

The study of matrices is based on the concept of linear transformations between two vector spaces. It is therefore necessary to define what this concept means in order to understand the setup of a matrix. In this chapter, as well as in the remainder of the book, the set of all real numbers is denoted by R, and its elements are referred to as scalars. The set of all n-tuples of real numbers will be denoted by R^n ($n \geq 1$).

1.1 VECTOR SPACES

This section introduces the reader to ideas that are used extensively in many books on linear and matrix algebra. They involve extensions of the Euclidean geometry which are important in the current mathematical literature and are described here as a convenient introductory reference for the reader. We confine ourselves to real numbers and to vectors whose elements are real numbers.

1.1.1 Euclidean Space

A vector $(x_0, y_0)'$ of two elements can be thought of as representing a point in a two-dimensional Euclidean space using the familiar Cartesian x, y coordinates, as in Figure 1.1. Similarly, a vector $(x_0, y_0, z_0)'$ of three elements can represent a point in a three-dimensional Euclidean space, also shown in Figure 1.1. In general, a vector of n elements can be said to represent a point (an n-tuple) in what is called an n-dimensional Euclidean space. This is a special case of a wider concept called a vector space, which we now define.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

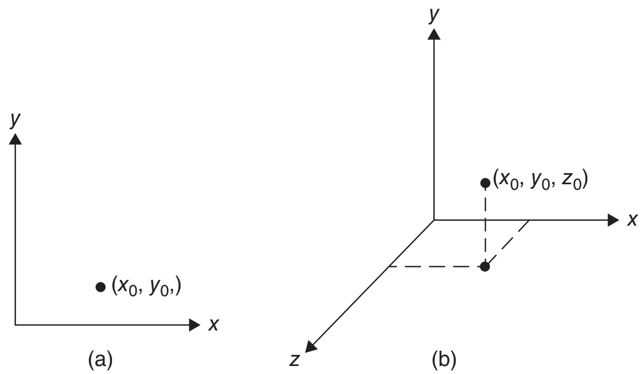


Figure 1.1 *(a) Two-Dimensional and (b) Three-Dimensional Euclidean Spaces.*

Definition 1.1 (Vector Spaces) *A vector space over R is a set of elements, denoted by V, which can be added or multiplied by scalars, in such a way that the sum of two elements of V is an element of V, and the product of an element of V by a scalar is an element of V. Furthermore, the following properties must be satisfied:*

(1) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ *for all* $\boldsymbol{u}$, $\boldsymbol{v}$ *in V.*

(2) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ *for all* $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ *in V.*

(3) *There exists an element in V, called the zero element and is denoted by* $\mathbf{0}$, *such that* $\mathbf{0} + \boldsymbol{u} = \boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ *for every* $\boldsymbol{u}$ *in V.*

(4) *For each* $\boldsymbol{u}$ *in V, there exists a unique element* $-\boldsymbol{u}$ *in V such that* $\boldsymbol{u} + (-\boldsymbol{u}) = (-\boldsymbol{u}) + \boldsymbol{u} = \mathbf{0}$.

(5) *For every* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in V and any scalar* α, $\alpha(\boldsymbol{u} + \boldsymbol{v}) = \alpha\boldsymbol{u} + \alpha\boldsymbol{v}$.

(6) $(\alpha + \beta)\boldsymbol{u} = \alpha\boldsymbol{u} + \beta\boldsymbol{u}$ *for any scalars* α *and* β *and any* $\boldsymbol{u}$ *in V.*

(7) $\alpha(\beta\boldsymbol{u}) = (\alpha\beta)\boldsymbol{u}$ *for any scalars* α *and* β *and any* $\boldsymbol{u}$ *in V.*

(8) *For every* $\boldsymbol{u}$ *in V,* $\mathbf{1}\boldsymbol{u} = \boldsymbol{u}$, *where* $\mathbf{1}$ *is the number one, and* $0\boldsymbol{u} = \mathbf{0}$, *where* 0 *is the number zero.*

Vector spaces were first defined by the Italian mathematician Giuseppe Peano in 1888.

Example 1.1 *The Euclidean space* R^n *is a vector space whose elements are of the form* $(x_1, x_2, \ldots, x_n)$, $n \geq 1$. *For every pair of elements in* R^n *their sum is in* R^n, *and so is the product of a scalar and any elements that is in* R^n. *It is easy to verify that properties (1) through (8) in Definition 1.1 are satisfied. The zero element is* $(0, 0, \ldots, 0)$.

Example 1.2 *The set of all polynomials in x of of degree n or less of the form* $\sum_{i=0}^{n} a_i x^i$, *where the* a_i*'s are scalars, is a vector space: the sum of any two such polynomials is a polynomial of the same form, and so is the product of a scalar with a polynomial. For the zero element,* $a_i = 0$ *for* $\forall$ *i.*

Example 1.3 *The set of all positive functions defined on the closed interval* $[-2, 2]$ *is not a vector space since multiplying any such function by a negative scalar produces a function that is not in that set.*

Definition 1.2 (Vector Subspace) *Let V be a vector space over R, and let W be a subset of V. Then W is said to be a vector subspace of V if it satisfies the following conditions:*

(1) *The sum of any two elements in W is an element of W.*
(2) *The product of any element in W by any scalar is an element in W.*
(3) *The zero element of V is also an element of W.*

It follows that for W to be a vector subspace of V, it must itself be a vector space. A vector subspace may consist of one element only, namely the zero element.

The set of all continuous functions defined on the closed interval $[a, b]$ is a vector subspace of all functions defined on the same interval. Also, the set of all points on the straight line $2x - 5y = 0$ is a vector subspace of R^2. However, the set of all points on any straight line in R^2 not going through the origin $(0, 0)$ is not a vector subspace.

Example 1.4 *Let V_1, V_2, and V_3 be the sets of vectors having the forms $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, respectively:*

$$\boldsymbol{x} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ \beta \end{bmatrix}, \quad \text{and} \quad \boldsymbol{z} = \begin{bmatrix} \gamma \\ 0 \\ \delta \end{bmatrix} \quad \text{for real } \alpha,\ \beta,\ \gamma, \text{ and } \delta.$$

Then V_1, V_2, and V_3 each define a vector space, and they are all subspaces of R^3. Furthermore, V_1 and V_2 are each a subspace of V_3.

1.2 BASE OF A VECTOR SPACE

Suppose that every element in a vector space V can be expressed as a linear combination of a number of elements in V. The set consisting of such elements is said to *span* or *generate* the vector space V and is therefore called a *spanning set* for V.

Definition 1.3 *Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ be elements in a vector space V. If there exist scalars $a_1, a_2, \dots, a_n$, not all equal to zero, such that $\sum_{i=1}^{n} a_i \boldsymbol{u}_i = \boldsymbol{0}$, then $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are said to be linearly dependent. If, however, $\sum_{i=1}^{n} a_i \boldsymbol{u}_i = \boldsymbol{0}$ is true only if all the a_i's are zero, then $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are said to be linearly independent.*

Note 1:
If $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are linearly independent, then none of them can be zero. To see this, if, for example, $\boldsymbol{u}_1 = \boldsymbol{0}$, then $a_1\boldsymbol{u}_1 + 0\boldsymbol{u}_2 + \cdots + 0\boldsymbol{u}_n = \boldsymbol{0}$ for any $a_1 \neq 0$, which implies that $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are linearly dependent, a contradiction. It follows that any set of elements of V that contains the zero element $\boldsymbol{0}$ must be linearly dependent. Furthermore, if $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are linearly dependent, then at least one of them can be expressed as a linear combination of the remaining elements.

Example 1.5 *Consider the vectors*

$$\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix}, \text{ and } \boldsymbol{x}_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}.$$

It is clear that

$$2\boldsymbol{x}_1 + \boldsymbol{x}_4 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}, \tag{1.1}$$

that is,

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_4 = \mathbf{0} \quad \text{for} \quad \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

which is not zero. Therefore, $\boldsymbol{x}_1$ and $\boldsymbol{x}_4$ are linearly dependent. So also are $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$ because

$$2\boldsymbol{x}_1 + 3\boldsymbol{x}_2 - 3\boldsymbol{x}_3 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} 0 \\ 15 \\ -15 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}. \tag{1.2}$$

In contrast to (1.1) and (1.2), consider

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 = \begin{bmatrix} 3a_1 \\ -6a_1 \\ 9a_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5a_2 \\ -5a_2 \end{bmatrix} = \begin{bmatrix} 3a_1 \\ -6a_1 + 5a_2 \\ 9a_1 - 5a_2 \end{bmatrix}. \tag{1.3}$$

There are no values a_1 and a_2 which make (1.3) a zero vector other than $a_1 = 0 = a_2$. Therefore $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent.

Definition 1.4 *If the elements of a spanning set for a vector space V are linearly independent, then the set is said to be a basis for V. The number of elements in this basis is called the dimension of V and is denoted by dim(V).*

Note 2:
The reference in Definition 1.4 was to *a* basis and not the basis because for any vector space there are many bases. All bases of a space V have the same number of elements which is $\dim(V)$.

Example 1.6 *The vectors,*

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \text{ and } \quad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

are all in R^3. Any two of them form a basis for the vector space whose typical vector is $(\alpha, \beta, 0)'$ for α and β real. The dimension of the space is 2. (The space in this case is, of course, a subspace of R^3.)

Note 3:
If $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ form a basis for V, and if $\boldsymbol{u}$ is a given element in V, then there exists a unique set of scalars, $a_1, a_2, \ldots, a_n$, such that $\boldsymbol{u} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i$. (see Exercise 1.4).

1.3 LINEAR TRANSFORMATIONS

Linear transformations concerning two vector spaces are certain functions that map one vector space, U, into another vector space, V. More specifically, we have the following definition:

Definition 1.5 *Let U and V be two vector spaces over R. Suppose that T is a function defined on U whose values belong to V, that is, T maps U into V. Then, T is said to be a linear transformation on U into V if*

$$T(a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2) = a_1 T(\boldsymbol{u}_1) + a_2 T(\boldsymbol{u}_2),$$

for all $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ in U and any scalars a_1, a_2.

For example, let T be a function from R^3 into R^3 such that $T(x_1, x_2, x_3) = (x_1 - x_2, x_1 + x_3, x_3)$. It can be verified that T is a linear transformation.

Example 1.7 *In genetics, the three possible genotypes concerning a single locus on a chromosome at which there are only two alleles, G and g, are GG, Gg, and gg. Denoting these by g_1, g_2, and g_3, respectively, gene effects relative to these genotypes can be defined (e.g., Anderson and Kempthorne, 1954) in terms of a mean μ, a measure of gene substitution α, and a measure of dominance δ, such that*

$$\begin{aligned}
\mu &= \frac{1}{4}g_1 + \frac{1}{2}g_2 + \frac{1}{4}g_3 \\
\alpha &= \frac{1}{4}g_1 \qquad\qquad - \frac{1}{4}g_3 \\
\delta &= -\frac{1}{4}g_1 + \frac{1}{2}g_2 - \frac{1}{4}g_3
\end{aligned}$$

These equations represent a linear transformation of the vector $(g_1, g_2, g_3)'$ to $(\mu, \alpha, \delta)'$. In Chapter 2, it will be seen that such a transformation is determined by an array of numbers consisting of the coefficients of the g_i's in the above equations. This array is called a *matrix*.

1.3.1 The Range and Null Spaces of a Linear Transformation

Let T be a linear transformation that maps U into V, where U and V are two given vector spaces. The image of U under T, also called the *range* of T and is denoted by $\mathfrak{R}(T)$, is the set of all elements in V of the form $T(\boldsymbol{u})$ for $\boldsymbol{u}$ in U. The *null space*, or the *kernel*, of T consists of all elements $\boldsymbol{u}$ in U such that $T(\boldsymbol{u}) = \mathbf{0}_v$ where $\mathbf{0}_v$ is the zero element in V. This space is denoted by $\aleph(T)$. It is easy to show that $\mathfrak{R}(T)$ is a vector subspace of V and $\aleph(T)$ is a vector subspace of U (see Exercise 1.7). For example, let T be a linear transformation from R^3 into R^2 such that $T(x_1, x_2, x_3) = (x_1 - x_3, x_2 - x_1)$, then $\aleph(T)$ consists of all points in R^3 such that $x_1 - x_3 = 0$ and $x_2 - x_1 = 0$, or equivalently, $x_1 = x_2 = x_3$. These equations represent a straight line in R^3 passing through the origin.

Theorem 1.1 *Let T be a linear transformation from the vector space U into the vector space V. Let $n = dim(U)$. Then, $n = p + q$, where $p = dim[\aleph(T)]$ and $q = dim[\mathfrak{R}(T)]$.*

Proof. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$ be a basis for $\aleph(T)$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_q$ be a basis for $\mathfrak{R}(T)$. Furthermore, let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ be elements in U such that $T(\boldsymbol{w}_i) = \boldsymbol{v}_i$ for $i = 1, 2, \ldots, q$. Then,

1. $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are linearly independent.
2. The elements in (1) span U.

To show (1), suppose that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are not linearly independent, then there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_p; \beta_1, \beta_2, \ldots, \beta_q$ such that

$$\sum_{i=1}^{p} \alpha_i \boldsymbol{u}_i + \sum_{i=1}^{q} \beta_i \boldsymbol{w}_i = \mathbf{0}_u, \tag{1.4}$$

where $\mathbf{0}_u$ is the zero element in U. Mapping both sides of (1.4) under T, we get

$$\sum_{i=1}^{p} \alpha_i T(\boldsymbol{u}_i) + \sum_{i=1}^{q} \beta_i T(\boldsymbol{w}_i) = \mathbf{0}_v,$$

where $\mathbf{0}_v$ is the zero elements in V. Since the $\boldsymbol{u}_i$'s belong to the null space, then

$$\sum_{i=1}^{q} \beta_i T(\boldsymbol{w}_i) = \sum_{i=1}^{q} \beta_i \boldsymbol{v}_i = \mathbf{0}_v.$$

But, the $\boldsymbol{v}_i$'s are linearly independent, therefore, $\beta_i = 0$ for $i = 1, 2, \ldots, q$. From (1.4) it can be concluded that $\alpha_i = 0$ for $i = 1, 2, \ldots, p$ since the $\boldsymbol{u}_i$'s are linearly independent. It follows that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are linearly independent.

To show (2), let $\boldsymbol{u}$ be any element in U, and let $T(\boldsymbol{u}) = \boldsymbol{v}$. There exist scalars $a_1, a_2, \ldots, a_q$ such that $\boldsymbol{v} = \sum_{i=1}^{q} a_i \boldsymbol{v}_i$. Hence,

$$\begin{aligned} T(\boldsymbol{u}) &= \sum_{i=1}^{q} a_i T(\boldsymbol{w}_i) \\ &= T\left(\sum_{i=1}^{q} a_i \boldsymbol{w}_i\right). \end{aligned}$$

It follows that

$$T\left(\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i\right) = \boldsymbol{0}_v.$$

This indicates that $\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i$ is an element in $\aleph(T)$. We can therefore write

$$\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i = \sum_{i=1}^{p} b_i \boldsymbol{u}_i, \tag{1.5}$$

for some scalars $b_1, b_2, \ldots, b_p$. From (1.5) it follows that $\boldsymbol{u}$ can be written as a linear combination of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$; $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$, which proves (2).

From (1) and (2) we conclude that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$; $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ form a basis for U. Hence, $n = p + q$. ■

REFERENCE

Anderson, V. L. and Kempthorne, O. (1954). A model for the study of quantitative inheritance. *Genetics*, 39, 883–898.

EXERCISES

1.1 For $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} -13 \\ -1 \\ 2 \end{bmatrix}$, and $\mathbf{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$,

show the following:

(a) $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are linearly dependent, and find a linear relationship among them.

(b) $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_4$ are linearly independent, and find the linear combination of them that equals $(a, b, c)'$.

1.2 Let U and V be two vector spaces over R. The *Cartesian product* $U \times V$ is defined as the set of all ordered pairs $(\boldsymbol{u}, \boldsymbol{v})$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are elements in U and V, respectively. The sum of two elements, $(\boldsymbol{u}_1, \boldsymbol{v}_1)$ and $(\boldsymbol{u}_2, \boldsymbol{v}_2)$ in $U \times V$ is defined as $(\boldsymbol{u}_1 + \boldsymbol{u}_2, \boldsymbol{v}_1 + \boldsymbol{v}_2)$, and if α is a scalar, then $\alpha(\boldsymbol{u}, \boldsymbol{v})$ is defined as $(\alpha\boldsymbol{u}, \alpha\boldsymbol{v})$, where $(\boldsymbol{u}, \boldsymbol{v})$ is an element in $U \times V$. Show that $U \times V$ is a vector space.

1.3 Let $\boldsymbol{V}$ be a vector space spanned by the vectors,

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \ \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 4 \\ 6 \end{bmatrix}, \text{ and } \boldsymbol{v}_4 = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

(a) Show that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are linearly independent.

(b) Show that $\boldsymbol{V}$ has dimension 2.

1.4 Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ be a basis for a vector space $\boldsymbol{U}$. Show that if $\boldsymbol{u}$ is any element in $\boldsymbol{U}$, then there exists a unique set of scalars, $a_1, a_2, \ldots, a_n$, such that $\boldsymbol{u} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i$, which proves the assertion in Note 3.

1.5 Let $\boldsymbol{U}$ be a vector subspace of $\boldsymbol{V}$, $\boldsymbol{U} \neq \boldsymbol{V}$. Show that $\dim(\boldsymbol{U}) < \dim(\boldsymbol{V})$.

1.6 Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$ be elements in a vector space $\boldsymbol{U}$. The collection of all linear combinations of the form $\sum_{i=1}^{m} a_i \boldsymbol{u}_i$, where $a_1, a_2, \ldots, a_m$ are scalars, is called the linear span of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$ and is denoted by $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$. Show that $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$ is a vector subspace of $\boldsymbol{U}$.

1.7 Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be vector spaces and let T be a linear transformation from $\boldsymbol{U}$ into $\boldsymbol{V}$.

1. Show that $\aleph(T)$, the null space of T, is a vector subspace of $\boldsymbol{U}$.

2. Show that $\Re(T)$, the range of T, is a vector subspace of $\boldsymbol{V}$.

1.8 Consider a vector subspace of R^4 consisting of all $\boldsymbol{x} = (x_1, x_2, x_3, x_4)'$ such that $x_1 + 3x_2 = 0$ and $2x_3 - 7x_4 = 0$. What is the dimension of this vector subspace?

1.9 Suppose that T is a linear transformation from R^3 onto R (the image of R^3 under T is all of R) given by $T(x_1, x_2, x_3) = 3x_1 - 4x_2 + 9x_3$. What is the dimension of its null space?

1.10 Let T be a linear transformation from the vector space U into the vector space V. Show that T is one-to-one if and only if whenever $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are linearly independent in U, then $T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_n)$ are linearly independent in V.

1.11 Let the functions x, e^x be defined on the closed interval $[0, 1]$. Show that these functions are linearly independent.

2

Matrix Notation and Terminology

We have seen in the Introduction that a matrix is defined in terms of a system of m linear equations in n variables of the form

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \quad i = 1, 2, \ldots, m, \tag{2.1}$$

where the a_{ij}'s are given coefficients. These equations can be represented by a single equation as shown in (2.2)

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_m \end{bmatrix}. \tag{2.2}$$

This representation is in fact a linear transformation from the n-dimensional Euclidean space R^n to the m-dimensional Euclidean space R^m, that is, $R^n \rightarrow R^m$. The array of a_{ij} coefficients in (2.2) is called a *matrix* with m rows and n columns.

Thus a matrix is a rectangular or square array of numbers arranged in rows and columns. The rows are of equal length, as are the columns. The a_{ij} coefficient denotes the element in

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

the ith row and jth column of a matrix $\boldsymbol{A}$. We then have the representation

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & & \vdots & & \vdots & \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & & \vdots & & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}. \tag{2.3}$$

The three dots indicate, in the first row, for example, that the elements a_{11}, a_{12}, and a_{13} continue in sequence up to a_{1j} and then up to a_{1n}; likewise in the first column the elements a_{11}, a_{21}, continue in sequence to a_{i1} and up to a_{m1}. The use of three dots to represent omitted values of a long sequence in this manner is standard and will be used extensively. This form of writing a matrix clearly specifies its elements, and also its size, that is, the number of rows and columns. An alternative and briefer form is

$$\mathbf{A} = \{a_{ij}\}, \quad \text{for} \quad i = 1, 2, \ldots m, \quad \text{and} \quad j = 1, 2, \ldots, n, \tag{2.4}$$

the curly brackets indicating that a_{ij} is a typical element, the limits of i and j being m and n, respectively.

The element a_{ij} is called the (i, j)th *element*, the first subscript referring to the row the element is in and the second to the column. Thus a_{23} is the element in the second row and third column. The size of the matrix, that is, the number of rows and columns, is referred to as its *order* or as its *dimensions*. (Sometimes the word *order* is also used for other characteristics of a matrix, but in this book it always refers to the number of rows and columns.) Thus $\mathbf{A}$ with m rows and n columns has order $m \times n$ (read as "m by n") and, to emphasize its order, the matrix can be written $\mathbf{A}_{m\times n}$. A simple example of a matrix of order 2×3 is

$$\mathbf{A}_{2\times3} = \begin{bmatrix} 4 & 0 & -3 \\ -7 & 2.73 & 1 \end{bmatrix}.$$

Notice that zero is legitimate as an element, that the elements need not all have the same sign, and that integers and decimal numbers can both be elements of the same matrix. The element in the first row and first column, a_{11}, the element in the upper left-hand corner of a matrix is called the *leading element* of the matrix.

When $r = c$, the number of rows equals the number of columns, $\mathbf{A}$ is square and is referred to as a *square matrix* and, provided there is no ambiguity, is described as being of order r. In this case the elements $a_{11}, a_{22}, a_{33}, \ldots, a_{rr}$ are referred to as the *diagonal elements* or *diagonal* of the matrix. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 7 & 0 & 7 \\ 5 & 2 & 9 & 1 \\ 0 & 3 & 4 & 6 \\ 8 & 0 & 5 & 7 \end{bmatrix}$$

is square of order 4 and its diagonal elements are 1, 2, 4, and 7. The elements of a square matrix that lie in a line parallel to and just below the diagonal are sometimes referred to as being on the *subdiagonal elements*; in the example they are 5, 3, 5. Elements of a square matrix other than the diagonal elements are called *off-diagonal* or *nondiagonal* elements.

A square matrix having zero for all its nondiagonal elements is described as a *diagonal matrix*; for example,

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -17 & 0 \\ 0 & 0 & 99 \end{bmatrix}$$

is a diagonal matrix. When $a_1, a_2, \ldots, a_n$ (some of which may be zero) are the diagonal elements of a diagonal matrix, useful notations for that matrix are

$$\begin{aligned} D\{a_1, a_2, \ldots, a_n\} &= \operatorname{diag}\{a_1, a_2, \ldots, a_n\} \\ &= D\{a_i\} = \operatorname{diag}\{a_i\} \quad \text{for} \quad i = 1, 2, \ldots, n. \end{aligned}$$

A square matrix with all elements above (or below) the diagonal being zero is called a *triangular matrix*. For example,

$$\mathbf{B} = \begin{bmatrix} 1 & 5 & 13 \\ 0 & -2 & 9 \\ 0 & 0 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 & 0 \\ 8 & 3 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

are triangular matrices. **B** is an *upper triangular matrix* and **C** is a *lower triangular matrix*.

Example 2.1 *Suppose an experienced taxicab driver has ascertained that when in Town 1 there is a probability of .2 that the next fare will be within Town 1 and a probability of .8 that it will be to Town 2. But when in Town 2 the probabilities are .4 of going to Town 1 and .6 of staying in Town 2. These probabilities can be assembled in a matrix*

$$\boldsymbol{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \{p_{ij}\} \quad \textit{for} \quad i, j = 1, 2. \tag{2.5}$$

Then p_{ij} is the probability when in Town i of the next fare being to Town j.

The matrix in (2.5) is an example of a transition probability matrix. It is an array of probabilities of making a transition from what is called state i to state j. (The states in the example are the two towns.) Thus p_{ij} is a transition probability, and hence the name: *transition probability matrix*.

In (2.5), the sum of the elements in each row (the *row sum*) is 1. This is so because whenever the taxi is in Town i it must, with its next fare, either stay in that town or go to the other one. This is a feature of all transition probability matrices. In general, for

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mm} \end{bmatrix} \\ &= \{p_{ij}\} \quad \text{for} \quad i, j = 1, 2, \ldots, m, \end{aligned} \tag{2.6}$$

corresponding to m different states, the sum of the probabilities of going from state i to anyone of the m states (including staying in state i) must be 1. Hence,

$$p_{i1} + p_{i2} + \cdots + p_{ij} + \cdots p_{im} = 1,$$

that is,

$$\sum_{j=1}^{m} p_{ij} = 1 \text{ for all } i = 1, 2, \ldots, m. \tag{2.7}$$

Situations like the taxicab example form a class of probability models known as *Markov chains*. They arise in a variety of ways: for example, in considering the probabilities of changes in the prime interest rate from week to week, or the probabilities of certain types of mating in genetics, or in studying the yearly mobility of labor where p_{ij} can be the probability of moving during a calendar year from category i in the labor force to category j. Assembling probabilities of this nature into a matrix then enables matrix algebra to be used as a tool for answering questions of interest; for example, the probabilities of moving from category i to category j in 2 years are given by $\mathbf{P}^2$, in 3 years by $\mathbf{P}^3$, and so on.

2.1 PLOTTING OF A MATRIX

There is an interesting method to provide a graphical representation of a matrix. Suppose that $\boldsymbol{M}$ is a matrix is of order $m \times n$. The elements in each column of $\boldsymbol{M}$ are plotted against the digits $1, 2, \ldots, m$ which label the rows of $\boldsymbol{M}$. Thus if m_{ij} is the (i, j)th element of $\boldsymbol{M}$, then for each j, $m_{1j}, m_{2j}, \ldots, m_{mj}$ are plotted against $1, 2, \ldots, m$, respectively. In other words, each matrix column is plotted as a separate broken line and the resulting n of such lines form a graphical representation of the matrix $\boldsymbol{M}$. This is demonstrated in Figure 2.1 for the matrix,

$$\boldsymbol{M} = \begin{bmatrix} 15 & 2 & 4 & 14 \\ 5 & 10 & 11 & 7 \\ 8 & 8 & 6 & 13 \\ 5 & 14 & 13 & 1 \end{bmatrix}. \tag{2.8}$$

Plot of the 4×4 identity matrix,

$$\boldsymbol{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.9}$$

is shown in Figure 2.2. More detailed discussions concerning plotting matrices will be given in Chapter 16.

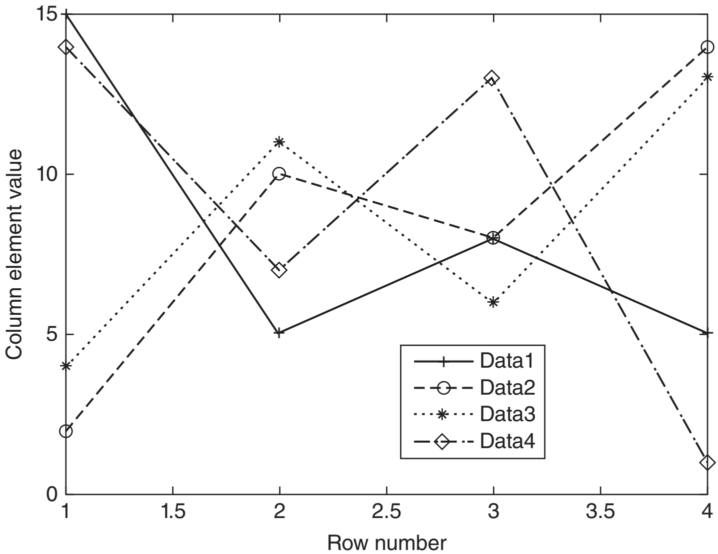


Figure 2.1 *Plot of Matrix* ***M****.*

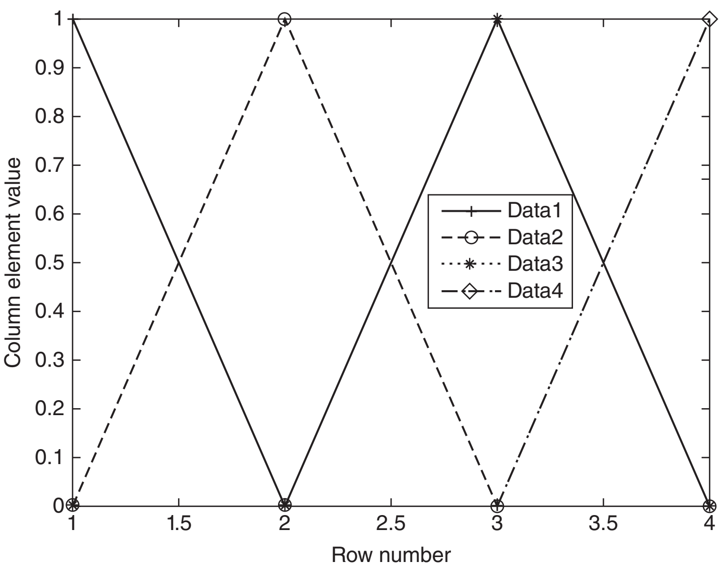


Figure 2.2 *Plot of a* 4×4 *Identity Matrix.*

2.2 VECTORS AND SCALARS

A matrix consisting of only a single column is called a *column vector*; for example,

$$\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

is a column vector of order 4. Likewise a matrix that is just a single row is a *row vector*:

$$\mathbf{y}' = [4 \quad 6 \quad -7]$$

is a row vector of order 3. A single number such as 2, 6.4 or –4 is called a *scalar*; the elements of a matrix are (usually) scalars. Sometimes it is convenient to think of a scalar as a matrix of order 1×1.

2.3 GENERAL NOTATION

A well-recognized notation is that of denoting matrices by uppercase letters and their elements by the lowercase counterparts with appropriate subscripts. Vectors are denoted by lowercase letters, often from the end of the alphabet, using the prime superscript to distinguish a row vector from a column vector. Thus $\mathbf{x}$ is a column vector and $\mathbf{x}'$ is a row vector. The lowercase Greek lambda, λ, is often used for a scalar.

Throughout this book the notation for displaying a matrix is

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{bmatrix},$$

enclosing the array of elements in square brackets. Among the variety of forms that can be found in the literature are

$$\begin{pmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{pmatrix}, \quad \begin{Bmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{Bmatrix}, \quad \text{and} \quad \begin{Vmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{Vmatrix}.$$

Single vertical lines are seldom used, since they are usually reserved for determinants (to be defined later). Another useful notation that has already been described is

$$\mathbf{A} = \{a_{ij}\} \quad \text{for } i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, c.$$

The curly brackets indicate that a_{ij} is a typical term of the matrix $\mathbf{A}$ for all pairs of values of i and j from unity up to the limits shown, in this case r and c; that is, $\mathbf{A}$ is a matrix of r rows and c columns. This is by no means a universal notation and several variants of it can be found in the literature. Furthermore, there is nothing sacrosanct about the repeated use in this book of the letter $\mathbf{A}$ for a matrix. Any letter may be used.

EXERCISES

2.1 For the elements of the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 17 & 9 & -2 & 3 \\ 3 & 13 & 10 & 2 & 6 \\ 11 & -9 & 0 & -3 & 2 \\ -6 & -8 & 1 & 4 & 5 \end{bmatrix} = \{a_{ij}\}$$

for $i = 1, 2, \ldots, 4$ and $j = 1, 2, \ldots, 5$, show that

(a) $a_{1.} = 26$;

(b) $a_{.3} = 20$;

(c) $\sum_{i=1}^{3} a_{i2} = 21$;

(d) $\sum_{\substack{i=1 \\ i \neq 2}}^{4} a_{i5} = 10$;

(e) $\sum_{i=3}^{4} a_{i.}^2 = \sum_{i=3}^{4} (a_{i.})^2 = 17$;

(f) $a_{..} = 57$;

(g) $\sum_{i=1}^{4} \frac{\sum_{j=1}^{5} a_{ij}^2}{a_{i.}} = 203\frac{275}{442}$;

(h) $\sum_{\substack{i=1 \\ i \neq 2}}^{4} \sum_{\substack{j=2 \\ j \neq 4}}^{5} a_{ij} = 20$;

(i) $\prod_{j=1}^{3} a_{4j} = \prod_{i=1}^{4} a_{i4}$;

(j) $\sum_{i=1}^{4} a_{i1} a_{i4} = -49$;

(k) $\sum_{i=1}^{4} \sum_{j=1}^{5} a_{..} = 1140$;

(l) $\prod_{i=1}^{3} a_{i4} = 12$;

(m) $\prod_{j=1}^{5} a_{2j} a_{3j} = 0$;

(n) $\prod_{i=1}^{4} 2^{a_{i4}} = 2$;

(o) $\sum_{\imath=1}^{4}\sum_{\jmath=1}^{5}(-1)^{\jmath}a_{\imath\jmath} = -29;$

(p) for $i = 2$, $\sum_{\substack{\jmath=1 \\ \jmath\neq 3}}^{5}(a_{\imath\jmath} - a_{\imath+2,\jmath})^2$
$= 527;$

(q) $\prod_{\jmath=1}^{5} 2^{(-1)^{j}a_{2\jmath}} = 0.0625.$

2.2 Write down the following matrices:

$$\begin{array}{llll} \mathbf{A} = \{a_{\imath\jmath}\} & \text{for} \quad a_{\imath\jmath} = i+j & \text{for} \quad i = 1, 2, 3 & \text{and} \quad j = 1, 2. \\ \mathbf{B} = \{b_{kt}\} & \text{for} \quad b_{kt} = k^{t-1} & \text{for} \quad k = 1, \ldots, 4 & \text{and} \quad t = 1, 3. \\ \mathbf{C} = \{c_{rs}\} & \text{for} \quad c_{rs} = 3r + 2(s-1) & \text{for} \quad r = 1, \ldots, 4 & \end{array}$$

and $s = 1, \ldots, 5$.

2.3 Prove the following identities:

(a)

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2.$$

(b)

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}\right)^2 = \sum_{i=1}^{m} a_{i.}^2.$$

(c)

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}a_{hk} = a_{i.}a_{h.}, \quad \textit{for} \quad i \neq h.$$

(d)

$$\sum_{i=1}^{m}\sum_{j=1}^{n} 4a_{ij} = 4a_{..}.$$

See Chapter 16 before doing Exercises 2.4 and 2.5.

2.4 Plot the following matrix:

$$\boldsymbol{A} = \begin{bmatrix} 13 & 4 & 6 & 16 \\ 7 & 8 & 11 & 9 \\ 9 & 10 & 8 & 15 \\ 6 & 12 & 15 & 2 \end{bmatrix}.$$

2.5 Plot the upper-triangular matrix,

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 7 \\ 0 & 7 & 3 \\ 0 & 0 & 9 \end{bmatrix}.$$

2.6 Construct a 3×3 matrix $\boldsymbol{A}$ whose (i, j)th element is $a_{ij} = (i + j)^2$. Conclude that the a_{ij}'s are symmetric with respect to the diagonal without doing any numerical computation.

2.7 Write down the matrices:

(a) $\boldsymbol{D}_1 = D\{3^{i-2}\}$ for $i = 1, 2, 3, 4$.

(b) $\boldsymbol{D}_2 = D\{i + 3^{i-2}\}$ for $i = 1, 2, 3, 4$.

2.8 If the elements a_{ij} of a matrix $\boldsymbol{A}$ of order $n \times n$ are symmetric with respect to its diagonal, how many different elements can $\boldsymbol{A}$ have above its diagonal?

2.9 Show that if the elements a_{ij} of a matrix $\boldsymbol{A}$ of order $n \times n$ are such that $a_{ij} = -a_{ji}$ for all $i, j(1, 2, \dots, n)$, then $\boldsymbol{A}$ can have in general $n(n-1)/2$ different elements above its diagonal, and the same identical elements (except for the negative sign) below the diagonal.

2.10 Suppose that

$$\begin{bmatrix} 2x & x - 2y \\ x - 3z & 3y + w \end{bmatrix} = \begin{bmatrix} 8 & 5 \\ 7 & 2 \end{bmatrix}.$$

Find the values of x, y, z, and w.

3

Determinants

In this chapter we discuss the calculation of a scalar value known as the *determinant* of a (square) matrix. Knowledge of this calculation is necessary for our discussion of matrix inversion and it is also useful in succeeding chapters for certain aspects of matrix theory proper that are used in applied problems. Moreover, since readers will undoubtedly encounter determinants elsewhere in their readings on matrices, it is appropriate to have an introduction to them here.

The literature of determinants is extensive and forms part of many texts on matrices. The presentation in this book is relatively brief, deals with elementary methods of evaluation and discusses selected topics arising therefrom. Extensive use is made of small numerical examples rather than rigorous mathematical proof. Several portions of the chapter can well be omitted at a first reading.

We begin with some general descriptions and proceed from them to the calculating of a determinant from its associated matrix. In practice, the method is useful mainly for deriving determinants of matrices of small order, but it is more informative about the structure of a determinant than is a formal definition. Nevertheless, a formal definition, which perforce is lengthy and quite difficult to follow, is given in Section 3.2.

3.1 EXPANSION BY MINORS

A determinant is a polynomial of the elements of a square matrix. It is a scalar. It is the sum of certain products of the elements of the matrix from which it is derived, each product being multiplied by +1 or −1 according to certain rules.

Determinants are defined only for square matrices—the determinant of a nonsquare matrix is undefined, and therefore does not exist. The determinant of a (square) matrix

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

of order $n \times n$ is referred to as an *n-order* determinant and the customary notation for the determinant of the matrix $\mathbf{A}$ is $|\mathbf{A}|$, it being assumed that $\mathbf{A}$ is square. Some texts use the notation $\|\mathbf{A}\|$ or det $\mathbf{A}$, but $|\mathbf{A}|$ is common and is used throughout this book. Obtaining the value of $|\mathbf{A}|$ by adding the appropriate products of the elements of A (with the correct + 1 or −1 factor included in the product) is variously referred to as *evaluating* the determinant, *expanding* the determinant or *reducing* the determinant. The procedure for evaluating a determinant is first illustrated by a series of simple numerical examples.

3.1.1 First- and Second-Order Determinants

To begin, we have the elementary result that the determinant of a 1×1 matrix is the value of its sole element.

The value of a second-order determinant is scarcely more difficult. For example, the determinant of

$$\mathbf{A} = \begin{bmatrix} 7 & 3 \\ 4 & 6 \end{bmatrix} \quad \text{is written} \quad |\mathbf{A}| = \begin{vmatrix} 7 & 3 \\ 4 & 6 \end{vmatrix}$$

and is calculated as

$$|\mathbf{A}| = 7(6) - 3(4) = 30.$$

This illustrates the general result for expanding a second-order determinant: from the product of the diagonal terms subtract the product of the off-diagonal terms. In other words, the determinant of a 2×2 matrix consists of the product (multiplied by +1) of the diagonal terms plus the product (multiplied by –1) of the off-diagonal terms. Hence, in general

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} + (-1)a_{12}a_{21} = a_{11}a_{22} - a_{12}a_{21}.$$

Further examples are

$$\begin{vmatrix} 6 & 8 \\ 17 & 21 \end{vmatrix} = 6(21) - 8(17) = -10$$

and

$$\begin{vmatrix} 9 & -3.81 \\ 7 & -1.05 \end{vmatrix} = 9(-1.05) - (-3.81)(7) = 17.22.$$

The evaluation of a second-order determinant is patently simple.

3.1.2 Third-Order Determinants

A third-order determinant can be expanded as a linear function of three second-order determinants derived from it. Their coefficients are elements of a row (or column) of the main determinant, each product being multiplied by $+1$ or -1. For example, the expansion of

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

based on the elements of the first row, 1, 2, and 3, is

$$\begin{aligned} |\mathbf{A}| &= 1(+1)\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 2(-1)\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} + 3(+1)\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1(50-48) - 2(40-42) + 3(32-35) = -3. \end{aligned}$$

The determinant that multiplies each element of the chosen row (in this case the first row) is the determinant derived from $|\mathbf{A}|$ by crossing out the row and column containing the element concerned. For example, the first element, 1, is multiplied by the determinant $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ which is obtained from $|\mathbf{A}|$ through crossing out the first row and column; and the element 2 is multiplied (apart from the factor of -1) by the determinant derived from $|\mathbf{A}|$ by deleting the row and column containing that element—namely, the first row and second column—so leaving $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$. Determinants obtained in this way are called *minors* of $|\mathbf{A}|$, that is to say, $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ is the minor of the element 1 in $|\mathbf{A}|$, and $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$ is the minor of the element 2.

The $(+1)$ and (-1) factors in the expansion are decided on according to the following rule: if $\mathbf{A}$ is written in the form $\mathbf{A} = \{a_{ij}\}$, the product of a_{ij} and its minor in the expansion of the determinant $|\mathbf{A}|$ is multiplied by $(-1)^{i+j}$. Hence because the element 1 in the example is the element a_{11}, its product with its minor is multiplied by $(-1)^{1+1} = +1$; and for the element 2, which is a_{12}, its product with its minor is multiplied by $(-1)^{1+2} = -1$.

Let us denote the minor of the element a_{11} by $|\mathbf{M}_{11}|$, where $\mathbf{M}_{11}$ is a submatrix of $\mathbf{A}$ obtained by deleting the first row and first column. Then in the preceding example $|\mathbf{M}_{11}| = \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$. Similarly, if $|\mathbf{M}_{12}|$ is the minor of a_{12}, then $|\mathbf{M}_{12}| = \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$; and if $|\mathbf{M}_{13}|$ is the minor of a_{13}, then $|\mathbf{M}_{13}| = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$. With this notation, the preceding expansion of $|\mathbf{A}|$ is

$$|\mathbf{A}| = a_{12}(-1)^{1+1}|\mathbf{M}_{11}| + a_{12}(-1)^{1+2}|\mathbf{M}_{12}| + a_{13}(-1)^{1+3}|\mathbf{M}_{13}|.$$

This method of expanding a determinant is known as *expansion by the elements of a row* (*or column*) or sometimes as *expansion by minors*. It has been illustrated using elements of the first row but can also be applied to the elements of any row (or column). For example, the expansion of $|\mathbf{A}|$ just considered, using elements of the second row gives

$$\begin{aligned} |\mathbf{A}| &= 4(-1)\begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 5(+1)\begin{vmatrix} 1 & 3 \\ 7 & 10 \end{vmatrix} + 6(-1)\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \\ &= -4(-4) + 5(-11) - 6(-6) = -3 \end{aligned}$$

as before, and using elements of the first column the expansion also gives the same result:

$$\begin{aligned}|\mathbf{A}| &= 1(+1)\begin{vmatrix}5 & 6\\ 8 & 10\end{vmatrix} + 4(-1)\begin{vmatrix}2 & 3\\ 8 & 10\end{vmatrix} + 7(+1)\begin{vmatrix}2 & 3\\ 5 & 6\end{vmatrix}\\ &= 1(2) - 4(-4) + 7(-3) = -3.\end{aligned}$$

The minors in these expansions are derived in exactly the same manner as in the first, by crossing out rows and columns of $|\mathbf{A}|$, and so are the $(+1)$ and (-1) factors; for example, the minor of the element 4 is $|\mathbf{A}|$ with second row and first column deleted, and since 4 is a_{21} its product with its minor is multiplied by $(-1)^{2+1} = -1$. Other terms are obtained in a similar manner.

The foregoing example illustrates the expansion of the general third-order determinant

$$|\mathbf{A}| = \begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix}.$$

Expanding this by elements of the first row gives

$$\begin{aligned}|\mathbf{A}| &= a_{11}(+1)\begin{vmatrix}a_{22} & a_{23}\\ a_{32} & a_{33}\end{vmatrix} + a_{12}(-1)\begin{vmatrix}a_{21} & a_{23}\\ a_{31} & a_{33}\end{vmatrix} + a_{13}(+1)\begin{vmatrix}a_{21} & a_{22}\\ a_{31} & a_{32}\end{vmatrix}\\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31}\\ &\qquad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

The reader should verify that expansion by the elements of any other row or column leads to the same result. No matter by what row or column the expansion is made, the value of the determinant is the same. Note that once a row or column is decided on and the sign calculated for the product of the first element therein with its minor, the signs for the succeeding products alternate from plus to minus and minus to plus.

The diagonal of a square matrix has already been defined. The *secondary diagonal* (as it is sometimes called) is the line of elements in a square matrix going from the lower left corner to the upper right, the "other" diagonal, so to speak. Thus the determinant of a matrix of order 2 is the product of elements on the diagonal minus that of the elements on the secondary diagonal. This extends to the determinant of a 3×3 matrix. From the end of the preceding paragraph, the expression for $|\mathbf{A}|$ can be rewritten with its three positive terms followed by its three negative terms:

$$\begin{aligned}|\mathbf{A}| = a_{11}a_{22}a_{33} &+ a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}\\ &- \left(a_{31}a_{22}a_{13} + a_{11}a_{32}a_{23} + a_{12}a_{21}a_{33}\right).\end{aligned}$$

3.1.3 *n*-Order Determinants

The expansion of an n-order determinant by this method is an extension of the expansion of a third-order determinant as just given. Thus the determinant of the $n \times n$ matrix $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, is obtained as follows. Consider the elements of any one row (or column): multiply each element, a_{ij}, of this row (or column) by its minor, $|\mathbf{M}_{ij}|$, the determinant derived from $|\mathbf{A}|$ by crossing out the row and column containing a_{ij}; multiply the product by $(-1)^{i+j}$; add the signed products and their sum is the determinant $|\mathbf{A}|$; that is, when expanding by elements of a row

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}| \quad \text{for any } i, \tag{3.1}$$

and when expanding by elements of a column

$$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}| \quad \text{for any } j. \tag{3.2}$$

This expansion is used recurrently when n is large, that is, each $|\mathbf{M}_{ij}|$ is expanded by the same procedure. Thus a fourth-order determinant is first expanded as four signed products each involving a third-order minor, and each of these is expanded as a sum of three signed products involving a second-order determinant. Consequently, a fourth-order determinant ultimately involves $4 \times 3 \times 2 = 24$ products of its elements, each product containing four elements. This leads us to the general statement that the determinant of a square matrix of order n is a sum *of* $(n!)^1$ signed products, each product involving n elements of the matrix. The determinant is referred to as an *n-order* determinant. Utilizing methods given by Aitken (1948), it can be shown that each product has one and only one element from each row and column and that all such products are included and none occurs more than once.

This method of evaluation requires lengthy computing for determinants of order exceeding 3 to 4, say, because for order n there are $n!$ terms to be calculated, and even for n as small as $n = 10$ this means 3,628,800 terms! Fortunately, easier methods exist, but because the method already discussed forms the basis of these easier methods, it has been considered first. Furthermore, it is useful in presenting various properties of determinants.

3.2 FORMAL DEFINITION

As has been indicated, the determinant of an n-order square matrix is the sum of $n!$ signed products of the elements of the matrix, each product containing one and only one element from every row and column of the matrix. Evaluation through expansion by elements of a row or column, as represented in equations formulas (3.1) and (3.2), yields the requisite products with their correct signs when the minors are successively expanded at each stage by the same procedure. We have already referred to this as the method of expansion by minors.

Example 3.1 *In the expansion of*

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

[1] $n!$, read as "n factorial" is the product of all integers 1 through n; e.g., $4! = 1(2)(3)(4) = 24$.

the j_1, j_2, j_3 array in the product $a_{12}a_{23}a_{31}$ is $j_1 = 2, j_2 = 3$, and $j_3 = 1$. Now n_1 is the number of j's in this array that are less than j_1 but follows it; that is, that are less than 2 but follow it. There is only one, namely $j_3 = 1$; therefore $n_1 = 1$. Likewise $n_2 = 1$ because $j_2 = 3$ is followed by only one j less than 3, and finally $n_3 = 0$. Thus

$$n_1 + n_2 + n_3 = 1 + 1 + 0 = 2$$

and the sign of the product $a_{12}a_{23}a_{31}$ in $|\mathbf{A}|$ is therefore $(-1)^2 = +1$. That the sign is also positive in the expansion of $|\mathbf{A}|$ by minors is easily shown, for in expanding by elements of the first row the term involving a_{12} is

$$-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -a_{12}(a_{21}a_{33} - a_{23}a_{31}),$$

which includes the term $+a_{12}a_{23}a_{31}$.

We shall now indicate the generality of the result implicit in the above example, that the formal definition of $|\mathbf{A}|$ agrees with the procedure of expansion by minors. The definition states that each product of n elements in $|\mathbf{A}|$ contains one element from every row and column. Suppose we set out to form one such product by selecting elements for it one at a time from each of the rows of $\mathbf{A}$. Starting with the first row there are n possible choices of an element as the first element of the product. Having made a choice there are then only $n - 1$ possible choices in the second row because the column containing the element chosen from the first row must be excluded from the choices in subsequent rows (in order to have one and only one element from each column as well as from each row). Similarly, the column containing the element chosen from the second row has to be excluded from the possible choices in the third and subsequent rows. Thus there are $(n - 2)$ possible choices in the third row, $(n - 3)$ in the fourth row, and so on. Hence, based on the definition, the total number of products in $|\mathbf{A}|$ is

$$n(n-1)(n-2)\cdots(3)(2)(1) = (n!).$$

On the other hand, expansion of $|\mathbf{A}|$ by minors initially gives $|\mathbf{A}|$ as a sum of n elements each multiplied (apart from sign) by its minor, a determinant of order $n - 1$. Each minor can be expanded in similar fashion as the sum of $(n - 1)$ elements multiplied (apart from sign) by minors that are determinants of order $n - 2$. In this way we see that the complete determinant consists of $n(n-1)(n-2)\cdots(3)(2)(1) = (n!)$ signed products each containing n elements. And at each stage of the expansion, the method of deleting a row and a column from a determinant to obtain the minor of an element ensures that in each product the n elements consist of one from every row and column of $\mathbf{A}$. Hence (apart from sign) both the definition and the procedure of expansion by minors lead to the same set of products.

We now show that the sign of each product is the same in both cases. The sign of any product in a determinant expanded by minors is the product of the signs applied to each minor involved in the derivation of the product. Consider the product as

$$a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n}.$$

Supposing the first expansion of $|\mathbf{A}|$ is by elements of the first row, the sign attached to the minor of a_{1J_1} is $(-1)^{1+J_1}$. If the expansion of this minor, is by elements of the second row of $|\mathbf{A}|$ which is now the first row of the minor, the minor therein of a_{2J_2} will have attached to it the sign $(-1)^{1+J_2-1}$ if j_1 is less than j_2 and $(-1)^{1+J_2}$ if j_1 exceeds j_2. Likewise the sign attached to the minor of a_{3J_3} in the minor of a_{2J_2} in the minor of a_{1J_1} in the expansion of $|\mathbf{A}|$ will be $(-1)^{1+J_2-2}$ if both j_1 and j_2 are less than j_3. It will be $(-1)^{1+J_2-1}$ if either of j_1 or j_2 is less than j_3 and the other exceeds it, and it will be $(-1)^{1+J_2}$ if both j_1 and j_2 exceed j_3. Hence the sign will be $(-1)^{1+J_3-m_3}$ where m_3 is the number of j's less than j_3 which precede it in the array $j_1j_2j_3 \cdots j_n$. In general the sign of the ith minor involved will be $(-1)^{1+j_l-m_l}$ where m_l is the number of j's less than j_i which precede it in the array. Therefore the combined sign is

$$\prod_{l=1}^{n} (-1)^{1+j_l-m_l} = (-1)^q \quad \text{for} \quad q = \sum_{i=1}^{n} (1 + j_l - m_i).$$

Now, since the j_l's are the first n integers in some order, the number of them that are less than some particular j_l is $j_i - 1$. Therefore, by the definitions of n_i and m_i their sum is $j_i - 1$, that is, $n_l + m_l = j_i - 1$, which gives

$$q = \sum_{i=1}^{n}(1 + j_l - m_l) = \sum_{l-1}^{n}(2 + n_l) = 2n + \sum_{l=1}^{n} n_l.$$

Therefore $(-1)^q = (-1)^p$ for $p = \sum_{l=1}^{n} n_l$, so that when the determinant is expanded by minors the sign of a product is $(-1)^p$ as specified by the definition. Hence the definition and the method of expansion by minors are equivalent.

3.3 BASIC PROPERTIES

Several basic properties of determinants are useful for circumventing both the tedious expansion by elements of a row (or column) described in Section 3.1 and the formal definition. Each is stated in the form of a theorem, followed by an example and proof.

3.3.1 Determinant of a Transpose

Theorem 3.1 *The determinant of the transpose of a matrix equals the determinant of the matrix itself:* $|\mathbf{A}'| = |\mathbf{A}|$.

Example 3.2 *On using expansion by the first row,*

$$|\mathbf{A}| = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 1 & 2 \\ 4 & 4 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 4 & 9 \end{vmatrix} + \begin{vmatrix} 2 & 2 \\ 4 & 9 \end{vmatrix} = 1 + 10 = 11$$

and

$$|\mathbf{A}'| = \begin{vmatrix} 1 & 2 & 4 \\ -1 & 1 & 4 \\ 0 & 2 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 4 \\ 2 & 9 \end{vmatrix} - 2\begin{vmatrix} -1 & 4 \\ 0 & 9 \end{vmatrix} + 4\begin{vmatrix} -1 & 1 \\ 0 & 2 \end{vmatrix}$$
$$= 1 + 18 - 8 = 11.$$

Proof. The formal definition of Section 3.2 and its equivalence to expansion by elements of a row provide the means for proving equality of that expansion with expansion by elements of a column. Therefore expansion of $|\mathbf{A}|$ by elements of its ith row is the same as expansion of $|\mathbf{A}'|$ by elements of its ith column, except that in $|\mathbf{A}|$ all minors will be of matrices that are the transpose of those in the corresponding minors in $|\mathbf{A}'|$. This will be true right down to the minors of the 2×2 matrices in each case. But these are equal, for example,

$$\begin{vmatrix} a & b \\ x & y \end{vmatrix} = ay - bx = \begin{vmatrix} a & x \\ b & y \end{vmatrix}.$$

Therefore, $|\mathbf{A}| = |\mathbf{A}'|$ ■

Corollary 3.1 *All properties of* $|\mathbf{A}|$ *in terms of rows can be stated equivalently in terms of columns—because expanding* $|\mathbf{A}'|$ *by rows is identical to expanding* $|\mathbf{A}|$ *by columns.*

3.3.2 Two Rows the Same

Theorem 3.2 *If two rows of* $\mathbf{A}$ *are the same,* $|\mathbf{A}| = 0$.

Example 3.3

$$\begin{vmatrix} 1 & 4 & 3 \\ 7 & 5 & 2 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 5 & 2 \\ 5 & 2 \end{vmatrix} - 4\begin{vmatrix} 7 & 2 \\ 7 & 2 \end{vmatrix} + 3\begin{vmatrix} 7 & 5 \\ 7 & 5 \end{vmatrix} = 0,$$

because

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0. \tag{3.3}$$

Proof. If $\mathbf{A}$ has two rows, the same, expand $|\mathbf{A}|$ by minors so that the 2×2 minors in the last step of the expansion are from two equal rows. Then (3.3) shows that all those minors are zero, and so $|\mathbf{A}| = 0$. ■

3.3.3 Cofactors

The signed minor $(-1)^{i+j}|\mathbf{M}_{ij}|$ used in formulas (3.1) and (3.2) is called a cofactor:

$$c_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|, \tag{3.4}$$

where $\mathbf{M}_{ij}$ is $\mathbf{A}$ with its ith row and jth column deleted.

Two properties of cofactors are important. The first is that the sum of products of the elements of a row with their own cofactors is the determinant:

$$\sum_J a_{\imath J}c_{\imath J} = \sum_j a_{ij}(-1)^{\imath+j}|\mathbf{M}_{\imath J}| = |\mathbf{A}| \quad \text{for each } i. \tag{3.5}$$

This is just formula (3.1). Second, the sum of products of elements of a row with cofactors of some other row is zero:

$$\sum_j a_{iJ}c_{kJ} = 0 \quad \text{for} \quad i \neq k. \tag{3.6}$$

This is so because, for $i \neq k$,

$$\sum_J a_{ij}c_{kj} = \sum_j a_{ij}(-1)^{k+J}|\mathbf{M}_{kj}| = 0$$

with the last equality being true because it represents, by comparison with (3.5), expansion of $|\mathbf{A}|$ with its kth row replaced by its ith row and so is a determinant having two rows the same and hence is zero. Formulas (3.5) and (3.6) can also be restated in terms of columns of $\mathbf{A}$.

Example 3.4 *In the determinant*

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

the cofactors of the elements of the first row are as follows:

$$\textit{that of the 1 is} \quad (-1)^{1+1}\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} = 50 - 48 = 2;$$

$$\textit{that of the 2 is} \quad (-1)^{1+2}\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} = -40 + 42 = 2;$$

$$\textit{that of the 3 is} \quad (-1)^{1+3}\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 32 - 35 = -3.$$

Multiplying these elements of the first row by their cofactors gives $|\mathbf{A}|$*:*

$$|\mathbf{A}| = 1(2) + 2(2) + 3(-3) = -3.$$

But multiplying the elements of another row, the second say, by these cofactors gives zero:

$$4(2) + 5(2) + 6(-3) = 0.$$

The determinantal form of this last expression is

$$4(-1)^{1+1}\begin{vmatrix}5 & 6\\ 8 & 10\end{vmatrix} + 5(-1)^{1+2}\begin{vmatrix}4 & 6\\ 7 & 10\end{vmatrix} + 6(-1)^{1+3}\begin{vmatrix}4 & 5\\ 7 & 8\end{vmatrix} = \begin{vmatrix}4 & 5 & 6\\ 4 & 5 & 6\\ 7 & 8 & 10\end{vmatrix},$$

which is clearly zero because two rows are the same.

3.3.4 Adding Multiples of a Row (Column) to a Row (Column)

Theorem 3.3 *Adding to one row (column) of a determinant any multiple of another row (column) does not affect the value of the determinant.*

Example 3.5

$$\begin{aligned}|\mathbf{A}| &= \begin{vmatrix}1 & 3 & 2\\ 8 & 17 & 21\\ 2 & 7 & 1\end{vmatrix}\\ &= 1\,(17-147) - 3\,(8-42) + 2\,(56-34) = 16.\end{aligned}$$

And adding four times row 1 to row 2 does not affect the value of $|\mathbf{A}|$*:*

$$\begin{aligned}|\mathbf{A}| &= \begin{vmatrix}1 & 3 & 2\\ 8+4 & 17+12 & 21+8\\ 2 & 7 & 1\end{vmatrix} = \begin{vmatrix}1 & 3 & 2\\ 12 & 29 & 29\\ 2 & 7 & 1\end{vmatrix}\\ &= 1\,(29-203) - 3\,(12-58) + 2\,(84-58)\\ &= -174 + 138 + 52 = 16.\end{aligned} \tag{3.7}$$

Proof. Suppose $\mathbf{B}$ of order $n \times n$ has $[b_{11}\ b_{12} \cdots b_{1n}]$ and $[b_{21}\ b_{22} \cdots b_{2n}]$ as its first two rows. Let $\mathbf{A}$ be $\mathbf{B}$ with λ times its second row added to its first. We show that $|\mathbf{A}| = |\mathbf{B}|$.

With $|\mathbf{M}_{ij}|$ being the minor of b_{ij} in $|\mathbf{B}|$, expansion of $|\mathbf{A}|$ by elements of its first row, using formula (3.1), gives

$$\begin{aligned}|\mathbf{A}| &= \sum_{j=1}^{n}(b_{1j} + \lambda b_{2j})(-1)^{1+j}|\mathbf{M}_{1j}|\\ &= \sum_{j=1}^{n} b_{1j}(-1)^{1+j}|\mathbf{M}_{1j}| + \lambda \sum_{j=1}^{n} b_{2j}(-1)^{1+j}|\mathbf{M}_{1j}|.\end{aligned}$$

The first sum here is $|\mathbf{B}|$, directly from (3.1). The second sum is an example of (3.6) and hence is zero. Therefore $|\mathbf{A}| = |\mathbf{B}|$ ■

3.3.5 Products

Theorem 3.4 $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ *when* $\mathbf{A}$ *and* $\mathbf{B}$ *are square and of the same order.*

Example 3.6 *With*

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 6 & 6 \end{bmatrix} = \begin{bmatrix} 16 & 15 \\ 60 & 57 \end{bmatrix}, |\mathbf{AB}| = 912 - 900 = 12,$$

and

$$|\mathbf{A}||\mathbf{B}| = \begin{vmatrix} 1 & 2 \\ 3 & 8 \end{vmatrix} \begin{vmatrix} 4 & 3 \\ 6 & 6 \end{vmatrix} = 2(6) = 12.$$

Before proving the theorem, we first show that any determinant can be reduced to the determinant of a triangular matrix, and then consider two useful lemmas.

Reduction to Triangular Form The determinant of a lower triangular matrix is the product of its diagonal elements. For example,

$$\begin{vmatrix} 6 & 0 & 0 \\ 3 & -1 & 0 \\ 7 & 3 & -5 \end{vmatrix} = 6\,(-1)\,(-5) = 30.$$

Verification is self-evident—through expansion by elements of successive rows.

Any determinant can be evaluated as the determinant of a lower triangular matrix. This is done by adding multiples of rows to other rows to reduce the determinant to triangular form. For example, consider

$$|\mathbf{P}| = \begin{vmatrix} 3 & 8 & 7 \\ 1 & 2 & 4 \\ -1 & 3 & 2 \end{vmatrix}.$$

We use row operations to reduce all elements above the diagonal to zero. Start first with the 2 of the (3, 3) element to reduce elements above it to zero; that is, to row 1 add (– 7/2) times row 3 and to row 2 add (−2) times row 3. This gives

$$|\mathbf{P}| = \begin{vmatrix} 6\frac{1}{2} & -2\frac{1}{2} & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix}.$$

Now use the (2, 2) element to reduce the elements above *it* to zero, by subtracting $-2\frac{1}{2}/(-4)$ times row 2 from row 1. Then

$$|\mathbf{P}| = \begin{vmatrix} 6\frac{1}{2} - 2\frac{1}{2}\,(3)\,/4 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix},$$

from which $|\mathbf{P}| = (37/8)(-4)(2) = -37$, as can be verified by any other form of expansion.

The method used in the example extends quite naturally to a determinant of any order n. Starting with the nth diagonal element, the elements above it are reduced to zero, then so are those above the $(n-1)$th,$(n-2)$th ,…,3rd, and 2nd diagonal elements. This yields a lower triangular matrix. (Similar calculations starting with the first diagonal element and using it to reduce all elements below it to zero, followed by doing the same to those below the 2nd, 3rd,…,$(n-1)$th, and nth diagonal elements yields an upper triangular matrix.) Then the determinant is the product of the resultant diagonal elements.

Two Useful Lemmas

Lemma 3.1 *For square matrices* $\mathbf{P}, \mathbf{Q}$, *and* $\mathbf{X}$ *of the same order* n

$$\begin{vmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{X} & \mathbf{Q} \end{vmatrix} = |\mathbf{P}||\mathbf{Q}|. \tag{3.8}$$

Proof. By row operations on the first n rows of the left-hand side of (3.8) reduce $\mathbf{P}$ to lower triangular form. Then expand the left-hand side by elements of successive rows for the first n rows. The result is $|\mathbf{P}||\mathbf{Q}|$, and so (3.8) is upheld. ∎

Example 3.7 *Using* $\mathbf{P}$ *of the preceding example and*

$$\mathbf{Q} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix},$$

we have

$$\begin{vmatrix} 3 & 8 & 7 & 0 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 & 0 & 0 & 0 \\ 3 & -4 & 0 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = |\mathbf{P}|\,|\mathbf{Q}|$$

$$= -37 \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}.$$

Lemma 3.2 *For* $\mathbf{R}$ *and* $\mathbf{S}$ *square and of the same order* n

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = |\mathbf{R}|. \tag{3.9}$$

Proof. Expand the left-hand side of (3.9) by elements of the successive columns through the $-\mathbf{I}$ to get

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = \left[(-1)^{n+1+1}(-1)\right]^n |\mathbf{R}| = (-1)^{n(n+3)}\,|\mathbf{R}| = |\mathbf{R}|.$$

∎

Determinant of a Product Consider the determinant of the matrix product

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix},$$

namely

$$\left| \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right| = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix}. \tag{3.10}$$

Note that using

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

in premultiplication in (3.10) merely adds multiples of rows of

$$\left| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right|$$

to other rows and so

$$\text{L. H. S. of (3.10)} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{B}|, \text{ by (3.8).}$$

But

$$\text{R. H. S. of (3.10)} = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{AB}|, \text{ by (3.9).}$$

Therefore $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$ or

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|. \tag{3.11}$$

Note that for all three determinants in (3.11) to exist, $\mathbf{A}$ and $\mathbf{B}$ must both be square, of the same order. This is also implied by (3.10). The result (3.11) does not hold when $\mathbf{A}$ or $\mathbf{B}$ or both are rectangular; even if $|\mathbf{AB}|$ exists in the form $|\mathbf{A}_{r\times c}\mathbf{B}_{c\times r}|$, $|\mathbf{A}|$ and $|\mathbf{B}|$ do not and (3.11) does not apply.

There are several useful corollaries to this result.

Corollary 3.2

1. $|\mathbf{AB}| = |\mathbf{BA}|$(*because* $|\mathbf{A}||\mathbf{B}| = |\mathbf{B}||\mathbf{A}|$).
2. $|\mathbf{A}^2| = |\mathbf{A}|^2$ *(each equals* $|\mathbf{A}||\mathbf{A}|$*);* $|\mathbf{A}^k| = |\mathbf{A}|^k$ *is the extension.*
3. *For orthogonal* $\mathbf{A}$, $|\mathbf{A}| = \pm 1$ *(because* $\mathbf{AA}' = \mathbf{I}$ *implies* $|\mathbf{A}|^2 = 1$*).*
4. *For idempotent* $\mathbf{A}$, $|\mathbf{A}| = 0$ *or 1 (because* $\mathbf{A}^2 = \mathbf{A}$ *implies* $|\mathbf{A}|^2 = |\mathbf{A}|$).

3.4 ELEMENTARY ROW OPERATIONS

In Section 3.3.4 we introduced the operation of adding a multiple of a row (column) of a determinant to another row (column). It did not alter the value of the determinant. That same operation on a matrix can be represented by a matrix product, for example,

$$\begin{bmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} \tag{3.12}$$

is the matrix product representation of adding 4 times row 1 to row 2, as used in (3.7). Applying the product rule of Section 3.3.5 gives

$$\begin{vmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix}$$

because

$$\begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1. \tag{3.13}$$

The matrix involved in (3.12) and (3.13)) is known as an *elementary operator matrix*. For the operation of adding λ times one row of a matrix to another, it is always an identity matrix with λ in an off-diagonal element; and its determinant is always unity. We denote such a matrix by $\mathbf{P}_{ij}(\lambda)$. Then $\mathbf{P}_{ij}(\lambda)\mathbf{A}$ is the operation on $\mathbf{A}$ of adding to its ith row λ times its jth row; and

$$\left|\mathbf{P}_{ij}(\lambda)\right| = 1; \quad \text{for example in (3.13)}, \quad \left|\mathbf{P}_{21}(4)\right| = 1. \tag{3.14}$$

Two other elementary operations are (i) interchanging two rows of a matrix, and (ii) multiplying a tow by a scalar. These are represented by matrices

$$\mathbf{E}_{ij} = \mathbf{I} \text{ with } i\text{th and } j\text{th rows interchanged}$$

and

$$\mathbf{R}_{ii}(\lambda) = \mathbf{I} \text{ with } i\text{th diagonal element replaced by } \lambda.$$

Then $\mathbf{E}_{ij}\mathbf{A}$ is $\mathbf{A}$ with its ith and jth rows interchanged, and $\mathbf{R}_{ii}(\lambda)\mathbf{A}$ is $\mathbf{A}$ with its ith row multiplied by λ

Example 3.8

$$\mathbf{E}_{12}\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 17 & 21 \\ 1 & 3 & 2 \\ 2 & 7 & 1 \end{bmatrix}$$

and

$$\mathbf{R}_{33}(5)\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 10 & 35 & 5 \end{bmatrix}.$$

All three of these elementary operator matrices have simple determinants:

$$|\mathbf{P}_{IJ}(\lambda)| = 1, \quad |\mathbf{R}_{Ii}(\lambda)| = \lambda \quad \text{and} \quad |\mathbf{E}_{IJ}| = -1. \tag{3.15}$$

Special cases of these elementary operator matrices used in combination with the product rule $|\mathbf{AB} = |\mathbf{A}||\mathbf{B}|$ *provide useful techniques for simplifying the expansion of a determinant.*

3.4.1 Factorization

Theorem 3.5 *When λ (a nonzero scalar) is a factor of a row (column) of* $|\mathbf{A}|$ *then it is also a factor of* $|\mathbf{A}|$*:*

$$|\mathbf{A}| = \lambda|\,\mathbf{A}\textit{ with } \lambda \textit{ factored out of a row (column)}|. \tag{3.16}$$

Example 3.9

$$\begin{vmatrix} 4 & 6 \\ 1 & 7 \end{vmatrix} = 2\begin{vmatrix} 2 & 3 \\ 1 & 7 \end{vmatrix}, \qquad \text{that is, } 4(7) - 1(6) = 22 = 2[2(7) - 1(3)].$$

Proof. When λ is a factor of every element in the ith row of $\mathbf{A}$,

$$\mathbf{A} = \mathbf{R}_{ii}(\lambda)[\mathbf{A} \text{ with } \lambda \text{ factored out of its } i\text{th row}]. \tag{3.17}$$

In the determinant of (3.17), use (3.11) and (3.15) to get (3.16). ■

Corollary 3.3 *If* $\mathbf{A}$ *is an $n \times n$ matrix and λ is a scalar, the determinant of the matrix* $\lambda\mathbf{A}$ *is* $\lambda^n|\mathbf{A}|$*; that is,* $|\lambda\mathbf{A}| = \lambda^n|\mathbf{A}|$*. (In this case λ is a factor of each of the n rows of* $\lambda\mathbf{A}$*, so that after factoring λ from each row the determinant* $|\mathbf{A}|$ *remains.)*

Example 3.10

$$\begin{bmatrix} 3 & 0 & 27 \\ -9 & 3 & 0 \\ 15 & 6 & -3 \end{bmatrix} = 3^3 \begin{bmatrix} 1 & 0 & 9 \\ -3 & 1 & 0 \\ 5 & 2 & -1 \end{bmatrix} = -2700.$$

Corollary 3.4 *If one row of a determinant is a multiple of another row, the determinant is zero. (Factoring out the multiple reduces the determinant to having two rows the same. Hence the determinant is zero.)*

Example 3.11

$$\begin{bmatrix} -3 & 6 & 12 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{bmatrix} = -(1.5)\begin{bmatrix} 2 & -4 & -8 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{bmatrix} = 0.$$

3.4.2 A Row (Column) of Zeros

Theorem 3.6 *When a determinant has zero for every element of a row (or column), the determinant is zero.*

Example 3.12

$$\begin{vmatrix} 0 & 0 & 0 \\ 3 & 6 & 5 \\ 2 & 9 & 7 \end{vmatrix} = 0\begin{vmatrix} 6 & 5 \\ 9 & 7 \end{vmatrix} - 0\begin{vmatrix} 3 & 5 \\ 2 & 7 \end{vmatrix} + 0\begin{vmatrix} 3 & 6 \\ 2 & 9 \end{vmatrix} = 0.$$

Proof. This is just (3.16) with $\lambda = 0$. ■

3.4.3 Interchanging Rows (Columns)

Theorem 3.7 *Interchanging two rows (columns) of a determinant changes its sign.*

Example 3.13 *Subtracting twice column 2 from column 1,*

$$\begin{bmatrix} 6 & 3 & 0 \\ -1 & 4 & 7 \\ 2 & 5 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 0 \\ -9 & 4 & 7 \\ -8 & 5 & 9 \end{bmatrix} = -3\,(-81 + 56) = 75,$$

whereas

$$\begin{bmatrix} 2 & 5 & 9 \\ -1 & 4 & 7 \\ 6 & 3 & 0 \end{bmatrix} = \begin{bmatrix} -8 & 5 & 9 \\ -9 & 4 & 7 \\ 0 & 3 & 0 \end{bmatrix} = -3\,(-56 + 81) = -75.$$

Proof. Using (3.11) and (3.15)

$$|\mathbf{E}_{ij}\mathbf{A}| = |\mathbf{E}_{ij}||\mathbf{A}| = -|\mathbf{A}|.$$

■

3.4.4 Adding a Row to a Multiple of a Row

In Section 3.3.4 we added a multiple of a row to a row. The reader is cautioned against doing the reverse: adding a row to a multiple of a row is not the same thing. It leads to a different result. For example, instead of adding λ times row 2 to row 1 of $|\mathbf{A}|$, and *not* altering $|\mathbf{A}|$, the operation of adding row 2 to λ times row 1 gives $\lambda|\mathbf{A}|$, for example,

$$\begin{vmatrix} 1 & \lambda \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2+5\lambda & 3+9\lambda \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = 3;$$

the operation which gives $\lambda|\mathbf{A}|$ is

$$\begin{vmatrix} \lambda & 1 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2\lambda+5 & 3\lambda+9 \\ 5 & 9 \end{vmatrix} = 3\lambda.$$

3.5 EXAMPLES

The foregoing properties can be applied in endless variation in expanding determinants. Efficiency in perceiving a procedure that leads to a minimal amount of effort in any particular case is largely a matter of practice, and beyond describing the possible steps available there is little more that can be said. The underlying method might be summarized as follows. By adding multiples of one row to other rows of the determinant results in reducing a column to having only one nonzero element. Expansion by elements of that column then involves only one minor, which is a determinant of order one less than the original determinant. Successive applications of this method reduce the determinant to one of order 2×2 whose expansion is obvious. If at any stage these reductions result in elements of a row containing a common factor, this can be factored out, and if they result in a row of zeros, or in two rows being identical, the determinant is zero.

Example 3.14

$$|\mathbf{A}| = \begin{vmatrix} 1 & 4 & 9 \\ -4 & 7 & 25 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 9 \\ -4+4(1) & 7+4(4) & 25+4(9) \\ 7-7(1) & 5-7(4) & 2-7(9) \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 4 & 9 \\ 0 & 23 & 61 \\ 0 & -23 & -61 \end{vmatrix} = 0.$$

The 1 in the leading element of $\mathbf{A}$ *makes it easy to reduce the other elements of the first column to zero, whereupon adding the last two rows gives* $|\mathbf{A}| = 0$.

Example 3.15

$$|\mathbf{B}| = \begin{vmatrix} x & y & y & y \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$

Observe that every column of $\mathbf{B}$ *sums to* $x + 3y$. *Therefore add every row of* $\mathbf{B}$ *to its first row and factor out* $x + 3y$*:*

$$|\mathbf{B}| = (x+3y)\begin{vmatrix} 1 & 1 & 1 & 1 \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$

Now use the leading element of $\mathbf{B}$ *to reduce the remainder of the first row to zeros. This is achieved by subtracting the first column from each of the other columns:*

$$|\mathbf{B}| = (x+3y)\begin{vmatrix} 1 & 0 & 0 & 0 \\ y & x-y & 0 & 0 \\ y & 0 & x-y & 0 \\ y & 0 & 0 & x-y \end{vmatrix} = (x+3y)\,(x-y)^3\,,$$

the last step being obtained by simple expansion by minors. It involves the more general result that the determinant of a diagonal matrix is the product of its diagonal elements:

$$\left|\mathbf{D}\left\{d_\iota\right\}\right| = \begin{vmatrix} d_1 & 0 & 0 & & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & \cdots & d_n \end{vmatrix} = \prod_{i=1}^{n} d_i = d_1 d_2 \cdots d_n.$$

Example 3.16

$$|\mathbf{C}| = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 18 & 27 & 3 & 13 & 5 \\ -11 & -17 & -2 & -9 & 2 \\ 7 & 10 & 1 & 4 & 7 \\ 4 & 3 & 13 & 8 & 5 \end{vmatrix} = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 0 & 3 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & 6 \\ 1 & 2 & 0 & 0 & 5 \\ -74 & -101 & 0 & -44 & -21 \end{vmatrix}.$$

The occurrence of a 1 as the (1, 3) element of $\mathbf{C}$ *prompts using it to reduce all other elements in column 3 to zero. Doing so, expansion by elements of the third column gives*

$$|\mathbf{C}| = (-1)^{1+3}\begin{vmatrix} 0 & 3 & 1 & -1 \\ 1 & -1 & -1 & 6 \\ 1 & 2 & 0 & 5 \\ -74 & -101 & -44 & -21 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 0 & 0 & -1 \\ 1 & 17 & 5 & 6 \\ 1 & 17 & 5 & 5 \\ -74 & -164 & -65 & -21 \end{vmatrix} = 0.$$

The (-1) *of the (1, 4) element has been used here to reduce other elements of the first row to zero; then expansion by elements of the first row gives a value of zero because the only nonzero element is associated with a* 3×3 *minor having two rows the same.*

Alternatively, in the original form of $|\mathbf{C}|$, *we can observe that row 2 plus row 3 equals row 4. Hence* $|\mathbf{C}| = 0$. *This is obviously easier—but this observation can be used only when it presents itself.*

Example 3.17

$$\begin{vmatrix} y & 7 & 7 & 7 \\ 7 & y & 7 & 7 \\ 7 & 7 & y & 7 \\ 7 & 7 & 7 & y \end{vmatrix} = 0 \textit{ for } y = 7 \textit{ or } -21$$

because for $y = 7$ *all rows are the same, and for* $y = -21$ *the sum of all rows is a row of zeros. This determinant is a special case of Example 3.15.*

Example 3.18

$$\begin{vmatrix} x & x & x \\ 4 & 3 & -9 \\ -3 & -2 & 10 \end{vmatrix} = 0 \quad \textit{for any } x,$$

because the first row, after factoring out x, equals the sum of the other two rows.

3.6 DIAGONAL EXPANSION

A matrix can always be expressed as the sum of two matrices one of which is a diagonal matrix, that is, as $(\mathbf{A} + \mathbf{D})$ where $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, and $\mathbf{D}$ is a diagonal matrix of order $n \times n$. The determinant of such a matrix can then be obtained as a polynomial of the elements of $\mathbf{D}$.

The minor of an element of a square matrix of order $n \times n$ is necessarily a determinant of order $n - 1$. But minors are not all of order $n - 1$. Deleting any r rows and any r columns from a square matrix of order $n \times n$ leaves a submatrix of order $(n - r) \times (n - r)$; the determinant of this submatrix is a *minor of order* $n - r$, or an $(n - r)$*-order minor*. It is useful to introduce an abbreviated notation for minors of the determinant of $\mathbf{A}$,

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}.$$

They will be denoted by just their diagonal elements; for example, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ is written as $\begin{vmatrix} a_{11} & a_{22} \end{vmatrix}$ and in similar fashion $\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$ is written as $\begin{vmatrix} a_{12} & a_{23} \end{vmatrix}$. Combined with the notation $\mathbf{A} = \{a_{ij}\}$ no confusion can arise. For example, $\begin{vmatrix} a_{21} & a_{32} \end{vmatrix}$ denotes the 2×2

minor having a_{21} and a_{32} as diagonal elements, and from $|\mathbf{A}|$ we see that the elements in the same rows and columns as these are a_{22} and a_{31}, so that

$$\left|a_{21} \quad a_{32}\right| = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Similarly,

$$\left|a_{21} \quad a_{33} \quad a_{44}\right| = \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix}.$$

We will now consider the determinant $|\mathbf{A}+\mathbf{D}|$, initially for a 2×2 case, denoting the diagonal elements of $\mathbf{D}$ by x_1 and x_2; for example,

$$|\mathbf{A}+\mathbf{D}| = \begin{vmatrix} a_{11}+x_1 & a_{12} \\ a_{21} & a_{22}+x_2 \end{vmatrix}.$$

By direct expansion

$$|\mathbf{A}+\mathbf{D}| = (a_{11}+x_1)(a_{22}+x_2) - a_{12}a_{21}.$$

Written as a function of x_1 and x_2 this is

$$x_1x_2 + x_1a_{22} + x_2a_{11} + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

In similar fashion it can be shown that

$$\begin{vmatrix} a_{11}+x_1 & a_{12} & a_{13} \\ a_{21} & a_{22}+x_2 & a_{23} \\ a_{31} & a_{32} & a_{33}+x_3 \end{vmatrix}$$
$$= x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11}$$
$$+x_1\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + x_2\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + x_3\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

which, using the abbreviated notation, can be written as

$$\begin{aligned} &x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11} \\ &\quad +x_1\left|a_{22} \quad a_{33}\right| + x_2\left|a_{11} \quad a_{33}\right| + x_3\left|a_{11} \quad a_{22}\right| \\ &\quad + \left|a_{11} \quad a_{22} \quad a_{33}\right|. \end{aligned} \tag{3.18}$$

Considered as a polynomial in the x's we see that the coefficient of the product of all the x's is unity; the coefficients of the second-degree terms in the x's are the diagonal elements of $\mathbf{A}$; the coefficients of the first-degree terms in the x's are the second-order minors of $|\mathbf{A}|$ having diagonals that are part (or all) of the diagonal of $|\mathbf{A}|$; and the term independent of the x's is $|\mathbf{A}|$ itself. The minors of $|\mathbf{A}|$ in these coefficients, namely those whose diagonals are coincident with the diagonal of $|\mathbf{A}|$, are called the *principal minors* of $|\mathbf{A}|$.

This method of expansion is useful on many occasions because the determinantal form $|\mathbf{A}+\mathbf{D}|$ occurs quite often, and when $|\mathbf{A}|$ is such that many of its principal minors are zero the expansion of $|\mathbf{A}+\mathbf{D}|$ by this method is greatly simplified.

Another way to introduce some of the above items is given by the following definition:

Definition 3.1 *If* $\mathbf{A}$ *is a square matrix of order* $n \times n$, *and if rows* $i_1, i_2, \ldots, i_r$ *and columns* $i_1, i_2, \ldots, i_r$ $(r < n)$ *are deleted from* $\mathbf{A}$, *then the resulting submatrix is called a principal submatrix of* $\mathbf{A}$ *(that is, this submatrix is obtained by deleting r rows and the same r columns from* $\mathbf{A}$*). The determinant of a principal submatrix is called a principal minor.*

If the deleted rows and columns are the last r rows and the last r columns, respectively, then the resulting submatrix is called a leading principal submatrix *of* $\mathbf{A}$. *Its determinant is called a leading principal minor.*

Example 3.19 *If*

$$|\mathbf{X}| = \begin{vmatrix} 7 & 2 & 2 \\ 2 & 8 & 2 \\ 2 & 2 & 9 \end{vmatrix},$$

we have

$$|\mathbf{X}| = |\mathbf{A}+\mathbf{D}| = \left\| \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} + \begin{bmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{bmatrix} \right\|.$$

Every element of $\mathbf{A}$ *is a 2 so that* $|\mathbf{A}|$ *and all its* 2×2 *minors are zero. Consequently* $|\mathbf{X}|$ *evaluated by (3.18) consists of only the first four terms:*

$$|\mathbf{X}| = 5\,(6)\,7 + 5\,(6)\,2 + 5\,(7)\,2 + 6\,(7)\,2 = 424.$$

Evaluating a determinant in this manner is also useful when all elements of the diagonal matrix $\mathbf{D}$ *are the same, that is, when the* x_i*'s are equal. The expansion (3.7) then becomes*

$$x^3 + x^2(a_{11} + a_{22} + a_{33}) + x(|a_{11} \quad a_{22}| + |a_{11}a_{33}| + |a_{22}a_{33}|) + |\mathbf{A}|,$$

which is generally written as

$$x^3 + x^2\mathrm{tr}_1(\mathbf{A}) + x\mathrm{tr}_2(\mathbf{A}) + |\mathbf{A}|$$

where $tr_1(\mathbf{A})$ *is the trace of* $\mathbf{A}$ *(sum of diagonal elements) and* $\mathrm{tr}_2(\mathbf{A})$ *is the sum of the principal minors of order 2 of* $|\mathbf{A}|$. *This method of expansion is known as expansion by diagonal elements or simply as diagonal expansion.*

The general diagonal expansion of a determinant of order n,

$$|\mathbf{A}+\mathbf{D}| = \begin{vmatrix} a_{11}+x_1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22}+x_2 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33}+x_3 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}+x_n \end{vmatrix},$$

consists of the sum of all possible products of the x_ι *taken* r *at a time for* $r = n, n-1, \dots, 2, 1, 0$, *each product being multiplied by its complementary principal minor of order* $n-r$ *in* $|\mathbf{A}|$. *By complementary principal minor in* $|\mathbf{A}|$ *is meant the principal minor having diagonal elements other than those associated in* $|\mathbf{A}+\mathbf{D}|$ *with the x's of the particular product concerned; for example, the complementary principal minor associated with* $x_1x_3x_6$ *is* $|a_{22}a_{44}a_{55}a_{77}a_{88}\cdots a_{nn}|$. *When the x's are all equal, the expression becomes* $\sum_{\iota=0}^{n} x^{n-i}\, tr_i(\mathbf{A})$ *where* $tr_i(\mathbf{A})$ *is the sum of the principal minors of order i of* $|\mathbf{A}|$ *and, by definition,* $\mathrm{tr}_0(\mathbf{A}) = 1$. *Note in passing that* $\mathrm{tr}_n(\mathbf{A}) = |\mathbf{A}|$.

Example 3.20 *Diagonal expansion gives*

$$|\mathbf{A}+\mathbf{D}| = \begin{vmatrix} a+b & a & a & a \\ a & a+b & a & a \\ a & a & a+b & a \\ a & a & a & a+b \end{vmatrix}$$
$$= b^4 + b^3 tr_1(\mathbf{A}) + b^2 tr_2(\mathbf{A}) + b tr_3(\mathbf{A}) + |\mathbf{A}|$$

where $\mathbf{A}$ *is the* 4×4 *matrix whose every element is a. Thus* $|\mathbf{A}|$ *and all minors of order 2 or more are zero. Hence*

$$|\mathbf{A}+\mathbf{D}| = b^4 + b^3(4a) = (4a+b)b^3.$$

This is an extension of the example in Section 3.4.

3.7 THE LAPLACE EXPANSION

In the expansion of

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

the minor of a_{11} is $|a_{22}a_{33}a_{44}|$. An extension of this, easily verified, is that the coefficient of $|a_{11} \quad a_{22}|$ is $|a_{33} \quad a_{44}|$; namely, the coefficient in $|\mathbf{A}|$ of

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \text{ is } \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} = a_{33}a_{44} - a_{34}a_{43}.$$

Likewise the coefficient of $|a_{11} \quad a_{24}|$ is $|a_{32} \quad a_{43}|$: that is, the coefficient of

$$|a_{11} \quad a_{24}| = \begin{vmatrix} a_{11} & a_{14} \\ a_{21} & a_{24} \end{vmatrix} = a_{11}a_{24} - a_{21}a_{14}$$

in the expansion of $|\mathbf{A}|$ is

$$\left| a_{32} \quad a_{43} \right| = \begin{vmatrix} a_{32} & a_{33} \\ a_{42} & a_{43} \end{vmatrix} = a_{32}a_{43} - a_{33}a_{42}.$$

Each determinant just described as the coefficient of a particular minor of $|\mathbf{A}|$ is the complementary minor in $|\mathbf{A}|$ of that particular minor: it is the determinant obtained from $|\mathbf{A}|$ by deleting from it all the rows and columns containing the particular minor. This is simply an extension of the procedure for finding the coefficient of an individual element in $|\mathbf{A}|$ as derived in the expansion by elements of a row or column discussed earlier. In that case, the particular minor is a single element and its coefficient in $|\mathbf{A}|$ is $|\mathbf{A}|$ amended by deletion of the row and column containing the element concerned. A sign factor is also involved, namely $(-1)^{i+j}$ for the coefficient of a_{ij} in $|\mathbf{A}|$. In the extension to coefficients of minors, the sign factor is minus one raised to the power of the sum of the subscripts of the diagonal elements of the chosen minor: for example, the sign factor for the coefficient of $\left| a_{32} \quad a_{43} \right|$ is $(-1)^{3+2+4+3} = -1$, as just given. The complementary minor multiplied by this sign factor can be appropriately referred to as the coefficient of the particular minor concerned. Furthermore, just as the expansion of a determinant is the sum of products of elements of a row (or column) with their coefficients, so also is the sum of products of all minors of order m that can be derived from any set of m rows, each multiplied by its coefficient as just defined. This generalization of the method of expanding a determinant by elements of a row to expanding it by minors of a set of rows was first established by Laplace and so bears his name. Aitken (1948) and Ferrar(1941) are two books where proof of the procedure is given; we shall be satisfied here with a general statement of the method and an example illustrating its use.

The Laplace expansion of a determinant $|\mathbf{A}|$ of order n can be obtained as follows. (i) Consider any m rows of $|\mathbf{A}|$. They contain $n!/\,[m!\,(n-m)!]$ minors of order m (see footnote, Section 3.1.) (ii) Multiply each of these minors, $\mathbf{M}$ say, by its complementary minor and by a sign factor, where (a) the complementary minor of $\mathbf{M}$ is the $n-m$ order minor derived from $|\mathbf{A}|$ by deleting the m rows and columns containing $\mathbf{M}$, and (b) the sign factor is $(-1)^{\mu}$ where μ is the sum of the subscripts of the diagonal elements of $\mathbf{M}$, $\mathbf{A}$ being defined as $\mathbf{A} = \{a_{ij}\}, i, j = 1, 2, \ldots, n$. (iii) The sum of all such products is $|\mathbf{A}|$.

Example 3.21 *Interchanging the second and fourth rows of*

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ 1 & 0 & 4 & 2 & 3 \\ 2 & 0 & 1 & 4 & 5 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix} \quad \textit{gives } |\mathbf{A}| = -\begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 2 & 0 & 1 & 4 & 5 \\ 1 & 0 & 4 & 2 & 3 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix}.$$

In this form we will expand $|\mathbf{A}|$ using the Laplace expansion based on the first two rows, $(m = 2)$. There are ten minors of order 2 in these two rows; seven of them are zero because *they involve a column of zeros. Hence* $|\mathbf{A}|$ *can be expanded as the sum of three products involving the three* 2×2 *nonzero minors in the first two rows, namely as* $-|\mathbf{A}| =$

$$(-1)^{1+1+2+2}\begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix}\begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 3 \\ 1 & 2 & 3 \end{vmatrix} + (-1)^{1+1+2+3}\begin{vmatrix} 1 & 3 \\ -1 & -1 \end{vmatrix}\begin{vmatrix} 0 & 4 & 5 \\ 0 & 2 & 3 \\ 2 & 2 & 3 \end{vmatrix}$$
$$+ (-1)^{1+2+2+3}\begin{vmatrix} 2 & 3 \\ 2 & -1 \end{vmatrix}\begin{vmatrix} 2 & 4 & 5 \\ 1 & 2 & 3 \\ 0 & 2 & 3 \end{vmatrix}.$$

The sign factors in these terms have been derived by envisaging $\mathbf{A}$ *as* $\{a_{ij}\}$. *Consequently the first* 2×2 *minor,* $\begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix}$, *is* $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, *leading to* $(-1)^{1+1+2+2}$ *as its sign factor; likewise for the other terms. Simplification of the whole expression gives*

$$-|\mathbf{A}| = 4\begin{vmatrix} 1 & 4 & 5 \\ 3 & 0 & 0 \\ 1 & 2 & 3 \end{vmatrix} - 2(2)\begin{vmatrix} 4 & 5 \\ 2 & 3 \end{vmatrix} + (-8)\begin{vmatrix} 2 & 4 & 5 \\ 1 & 0 & 0 \\ 0 & 2 & 3 \end{vmatrix} = -24 - 8 + 16$$

and hence $|\mathbf{A}| = 16$. *It will be found that expansion by a more direct method leads to the same result.*

Numerous other methods of expanding determinants are based on extensions of the Laplace expansion, using it recurrently to expand a determinant not only by minors and their complementary minors but also to expand these minors themselves. Many of these expansions are identified by the names of their originators, for example, Cauchy, Binet–Cauchy, and Jacoby. A good account of some of them is to be found in Aitken (1948) and Ferrar (1941)

3.8 SUMS AND DIFFERENCES OF DETERMINANTS

The sum of the determinants of each of two (or more) matrices generally does not equal the determinant of the sum. The simplest demonstration of this is

$$\begin{aligned} |\mathbf{A}| + |\mathbf{B}| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \\ &= a_{11}a_{22} - a_{12}a_{21} + b_{11}b_{22} - b_{12}b_{21} \neq |\mathbf{A} + \mathbf{B}|. \end{aligned}$$

The same applies to the difference, $|\mathbf{A}| - |\mathbf{B}| \neq |\mathbf{A} - \mathbf{B}|$.

We may note in passing that both $|\mathbf{A}| + |\mathbf{B}|$ and $|\mathbf{A}| - |\mathbf{B}|$ have meaning even when $\mathbf{A}$ and $\mathbf{B}$ are square matrices of different orders, because the value of a determinant is a scalar. This contrasts with $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$, which have meaning only when the matrices are conformable for addition (have the same order).

Another point of interest is that although $|\mathbf{A}| + |\mathbf{B}|$ does not generally equal $|\mathbf{A} + \mathbf{B}|$ the latter can be written as the sum of certain other determinants. For example,

$$\begin{aligned}
|\mathbf{A}+\mathbf{B}| &= \left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\right| = \begin{vmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}.
\end{aligned}$$

In general if $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$, $|\mathbf{A} + \mathbf{B}|$ can be expanded as the sum of 2^n n-order determinants.

Example 3.22 *The reader should verify that*

$$\begin{vmatrix} a+b & a & a \\ a & a+b & a \\ a & a & a+b \end{vmatrix} = (3a+b)b^2,$$

by expanding it as $|b\mathbf{I} + a\mathbf{J}|$.

3.9 A GRAPHICAL REPRESENTATION OF A 3 × 3 DETERMINANT

A 3×3 determinant can be depicted using a parallelpiped in a three-dimensional Euclidean space. To demonstrate this, the following definition is needed:

Definition 3.2 *Let* $\boldsymbol{u} = (u_1, u_2.u_3)'$, $\boldsymbol{v} = (v_1, v_2, v_3)'$, $\boldsymbol{w} = (w_1, w_2, w_3)'$ *be three vectors in a three-dimensional Euclidean space. Then,*

Dot Product *The* dot product *of two vectors such as* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, *denoted by* $\boldsymbol{u}.\boldsymbol{v}$, *is a scalar quantity equal to*

$$\boldsymbol{u}.\boldsymbol{v} = \| \boldsymbol{u} \| \| \boldsymbol{v} \| \, cos(\theta), \tag{3.19}$$

where θ *is the angle between the directions of* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, *and* $\| \boldsymbol{u} \|$ *and* $\| \boldsymbol{v} \|$ *are the Euclidean norms of* $\boldsymbol{u}$ *and* $\boldsymbol{v}$.
The dot product is also known as the scalar product. *It can be shown that* $\boldsymbol{u}.\boldsymbol{v}$ *is also equal to*

$$\boldsymbol{u}.\boldsymbol{v} = u_1 v_1 + u_2 v_2 + u_3 v_3. \tag{3.20}$$

Vector Product *The* vector product *of two vectors such as* $\boldsymbol{v}$ *and* $\boldsymbol{w}$, *denoted by* $\boldsymbol{v} \times \boldsymbol{w}$, *is a vector perpendicular to both* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *with the magnitude,*

$$\| \boldsymbol{v} \times \boldsymbol{w} \| = \| \boldsymbol{v} \| \| \boldsymbol{w} \| \, sin(\phi), \tag{3.21}$$

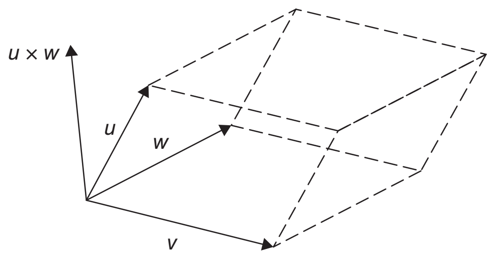


Figure 3.1 *Parallelpiped Constructed from the Vectors* **u, v, w**.

where ϕ is the angle between the directions of **v** *and* **w**. *The vector* $\mathbf{v} \times \mathbf{w}$ *is directed so that a rotation about it through the angle ϕ of not more than* 180° *carries* **v** *into* **w**. *It is known that* $\mathbf{v} \times \mathbf{w}$ *can be expressed as*

$$\boldsymbol{v} \times \boldsymbol{w} = (v_2w_3 - v_3w_2)\boldsymbol{e}_1 + (v_3w_1 - v_1w_3)\boldsymbol{e}_2 + (v_1w_2 - v_2w_1)\boldsymbol{e}_3, \quad (3.22)$$

where $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, *and* $\boldsymbol{e}_3$ *are unit vectors in the positive directions of the coordinate axes, x, y, z, respectively [see, e.g., Rutherford (1957, p. 7)].*

Triple Scalar Product *The* triple scalar product *of the three vectors* **u**, **v**, **w** *is a scalar quantity equal to* $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$, *that is, the dot product of the two vectors,* **u** *and* $\mathbf{v} \times \mathbf{w}$. *Hence, it is equal to*

$$\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w}) = \| \boldsymbol{u} \| \| \boldsymbol{v} \| \| \boldsymbol{w} \| \, sin(\phi)cos(\eta), \quad (3.23)$$

where η is the angle between **u** *and* $\mathbf{v} \times \mathbf{w}$.

It can be seen from (3.23) that the value of the triple scalar product $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$ is equal to the volume of the parallelpiped constructed from the vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$. This is true since $\| \boldsymbol{v} \| \| \boldsymbol{w} \| \sin(\phi)$ is the area of the base defined by $\boldsymbol{v}$ and $\boldsymbol{w}$, and $\| \boldsymbol{u} \| \cos(\eta)$ is the perpendicular height of the parallelpiped (see Figure 3.1). From (3.20) and (3.22), it follows that the triple product $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$ can be expressed as the value of the following determinant;

$$\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \quad (3.24)$$

Thus we have a graphical depiction of the determinant in (3.24) as the volume of the parallelpiped in Figure 3.1 which is constructed from the vectors that make up the rows of the determinant.

REFERENCES

Aitken, A. C. (1948). *Determinants and Matrices*, 5th ed. Oliver and Boyd, Edinburgh, UK.

Ferrar, W. L. (1941). *Algebra, A Text-Book of Determinants, Matrices and Algebraic Forms*. Oxford University Press, Oxford.

Rutherford, D. E. (1957). *Vector Methods*, 9th ed. Oliver and Boyd, Edinburgh, UK.

EXERCISES

3.1 Show that both

(a) $\begin{vmatrix} 1 & 5 & -5 \\ 3 & 2 & -5 \\ 6 & -2 & -5 \end{vmatrix}$ and $\begin{vmatrix} -3 & 2 & -6 \\ -3 & 5 & -7 \\ -2 & 3 & -4 \end{vmatrix}$ equal – 5;

(b) $\begin{vmatrix} 2 & 6 & 5 \\ -2 & 7 & -5 \\ 2 & -7 & 9 \end{vmatrix}$ and $\begin{vmatrix} 2 & -1 & 9 \\ -1 & 7 & 2 \\ 3 & -21 & 2 \end{vmatrix}$ equal 104.

3.2 **(a)** Show that the determinant of $\mathbf{x1}'$ is zero.

(b) If $\mathbf{A1} = \mathbf{0}$, why does $\mathbf{A}$ have a zero determinant?

3.3 For

$$\mathbf{A} = \begin{bmatrix} 0 & -a & b & -c \\ a & 0 & -d & e \\ -b & d & 0 & -f \\ c & -e & f & 0 \end{bmatrix}$$

(a) Calculate $|\mathbf{A}|$.

(b) Verify that $|\mathbf{I} + \mathbf{A}| = 1 + (a^2 + b^2 + c^2 + d^2 + e^2 + f^2) + |\mathbf{A}|$ by direct expansion

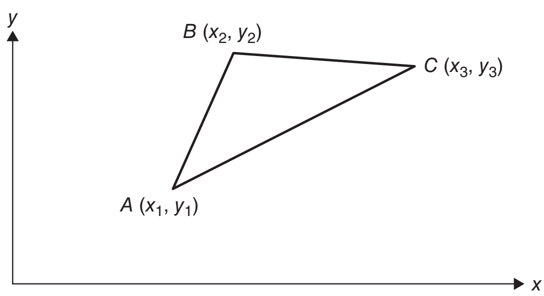
3.4 Without expanding the determinants, suggest values of x that satisfy the following equations:

(a) $\begin{vmatrix} x & x & x \\ 2 & -1 & 0 \\ 7 & 4 & 5 \end{vmatrix} = 0;$

(b) $\begin{vmatrix} 1 & x & x^2 \\ 1 & 2 & 4 \\ 1 & -1 & 1 \end{vmatrix} = 0;$

3.5 Drop perpendiculars from A, B, C in the accompanying figure and by using the resulting trapezoids show that the area of the triangle ABC can be expressed as the absolute value of $\frac{1}{2}|\mathbf{M}|$ where

$$\mathbf{M} = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$



3.6 The equation of a line in two co-ordinates can be written as $1 + ax + by = 0$. Show that the equation of a line passing through (x_1, y_1) and (x_2, y_2) is

$$\begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix} = 0.$$

3.7 Denote the equations

$$\left.\begin{aligned} x_1 + 2x_2 + 3x_3 &= 26 \\ 3x_1 + 7x_2 + 10x_3 &= 87 \\ 2x_1 + 11x_2 + 7x_3 &= 73 \end{aligned}\right\} \quad \text{by} \quad \mathbf{Ax} = \mathbf{b}.$$

(a) Solve the equation using successive elimination.

(b) Replace each column of $\mathbf{A}$, in turn, by $\mathbf{b}$. In replacing column j, call the resulting matrix $\mathbf{A}_j$. Verify that the solutions in (a) are

$$x_j = |\mathbf{A}_j|/|\mathbf{A}|.$$

This is known as Cramer's rule for solving linear equations.

(c) What condition must $|\mathbf{A}|$ satisfy for Cramer's rule to be workable?

3.8 Use the result in Section 3.3.4 to show that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b - a)(c - a)(c - b).$$

The above determinant is known as *Vandermonde determinant* of order 3.

3.9 As in Exercise 8, use the result in Section 3.3.4 to show that the Vandermonde determinant is equal to the determinant,

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ bc & ac & ab \end{vmatrix}$$

3.10 Let $A = (a_{ij})$ be a square matrix of order $n \times n$. If $\sum_j^n a_{ij} = 1$ for $i = 1, 2, \ldots, n$, show that $|A - I| = 0$.

3.11 Expand the determinant on the left using its first row and show that

$$\begin{vmatrix} a_1 + \lambda\alpha_1 & b_1 + \lambda\beta_1 & c_1 + \lambda\gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \lambda \begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

3.12 Show that a determinant is unaltered in value when to any row , or column, is added a constant multiple of any other row, or column.

(Hint: In Exercise 11, replace the second matrix on the right by a matrix whose second and third rows are identical to those in the matrix on the left, and its first row is equal to either the second row or the third row of the matrix on the left.)

3.13 Let $\boldsymbol{K}$ be a square matrix of order 12×12 such that $\boldsymbol{K}^5 = 3\boldsymbol{K}$. What is the numerical value of $|\boldsymbol{K}|$? Is there more than one possibility? If so, give all possible values.

3.14 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of order $m \times n$ and $n \times m$ $(n \geq m)$, respectively. Show that $|\boldsymbol{I}_m - \boldsymbol{AB}| = |\boldsymbol{I}_n - \boldsymbol{BA}|$.

3.15 Let $\boldsymbol{a}$ be a column vector of n elements. Show that

$$|\lambda \boldsymbol{I}_n + \boldsymbol{1}_n \boldsymbol{a}'| = \lambda^{n-1}(\lambda + \boldsymbol{a}'\boldsymbol{1}_n).$$

(Hint: Use the result in Exercise 14.)

3.16 Show that $|a\boldsymbol{I}_n + b\boldsymbol{J}_n| = a^{n-1}(a + nb)$, where $\boldsymbol{J}_n$ is a matrix of ones of order $n \times n$.

4

Matrix Operations

This chapter describes simple operations on matrices arising from their being rectangular arrays of numbers, and also the arithmetic of matrix addition, subtraction, and multiplication. Particular types of sums and products of, namely, the direct sum and direct product of matrices are defined. The trace, rank, and inverse of a matrix are also defined.

4.1 THE TRANSPOSE OF A MATRIX

Consider the following two matrices,

$$\mathbf{A} = \begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \\ 6 & 14 & 9 \\ 9 & 11 & 4 \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} 18 & 19 & 6 & 9 \\ 17 & 13 & 14 & 11 \\ 11 & 6 & 9 & 4 \end{bmatrix}.$$

Although the elements of **B** are the same as those of **A**, the matrices are clearly different; for example, they are not of the same order. **A** has order 4×3 and **B** is 3×4. The matrices are nevertheless related to each other, through the rows of one being the columns of the other. Whenever matrices are related in this fashion each is said to be the *transpose* of the other; for example, **B** is said to be the transpose of **A**, and **A** is the transpose of **B**.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

A formal description is that the transpose of a matrix $\mathbf{A}$ is the matrix whose columns are the rows of $\mathbf{A}$, with order retained, from first to last. The transpose is written as $\mathbf{A}'$, although the notations $\mathbf{A}^t$ and $\mathbf{A}^T$ are also seen in the literature. We will use $\mathbf{A}'$.

An obvious consequence of the transpose operation is that the rows of $\mathbf{A}'$ are the same as the columns of $\mathbf{A}$, and if $\mathbf{A}$ is $r \times c$, the order of $\mathbf{A}'$ is $c \times r$. If a_{ij} is the term in the ith row and the jth column of $\mathbf{A}$, it is also the term in the jth row and ith column of $\mathbf{A}'$. Therefore, on defining a'_{ji} as the element in the jth row and ith column of $\mathbf{A}'$, we have $a'_{ji} = a_{ij}$. Equivalently, on interchanging i and j, we have

$$\mathbf{A}' = \{a'_{ij}\} \quad \text{for} \quad i = 1, \ldots, c \quad \text{and} \quad j = 1, \ldots, r, \quad \text{and} \quad a'_{ij} = a_{ji}.$$

A minor difficulty in notation arises when, for example, $\mathbf{A}_{r\times c}$ has r rows and c columns; but combining the notation $\mathbf{A}'$ for the transpose of $\mathbf{A}$ with the $r \times c$ subscript notation for the order of $\mathbf{A}$, to get $\mathbf{A}'_{r\times c}$, is ambiguous. $\mathbf{A}'_{r\times c}$ could mean either that $\mathbf{A}$ of order $r \times c$ has been transposed or that the transpose of $\mathbf{A}$ has order $r \times c$ (and so $\mathbf{A}$ has order $c \times r$). For clarity, one of the equivalent forms $(\mathbf{A}_{r\times c})'$ or $(\mathbf{A}')_{c\times r}$ must be used whenever it is necessary to have subscript notation for the order of a transposed matrix. Fortunately, the need for this clumsy notation seldom arises. The simple $\mathbf{A}'$ suffices in most contexts.

Two important consequences of the transpose operation are worth noting.

4.1.1 A Reflexive Operation

The transpose operation is reflexive: the transpose of a transposed matrix is the matrix itself, that is, $(\mathbf{A}')' = \mathbf{A}$. This is so because transposing $\mathbf{A}'$ yields a matrix whose rows are the columns of $\mathbf{A}'$, and these are the rows of $\mathbf{A}$. Hence $(\mathbf{A}')' = \mathbf{A}$. More formally,

$$(\mathbf{A}')' = \{a'_{ij}\}' = \{a_{ji}\}' = \{a'_{ji}\} = \{a_{ij}\} = \mathbf{A}.$$

4.1.2 Vectors

The transpose of a row vector is a column vector and vice versa. For example, the transpose of

$$\mathbf{x} = \begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix} \qquad \text{is} \quad \mathbf{x}' = \begin{bmatrix} 1 & 6 & 4 \end{bmatrix}.$$

4.2 PARTITIONED MATRICES

4.2.1 Example

Consider the matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{bmatrix}.$$

Suppose we draw dashed lines between certain rows and columns as in

$$\mathbf{B} = \left[\begin{array}{cccc|cc} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ \hline 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{array}\right]. \tag{4.1}$$

Each of the arrays of numbers in the four sections of **B** engendered by the dashed lines is a matrix:

$$\mathbf{B}_{11} = \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \\ 4 & 3 & 6 & 1 \end{bmatrix}, \qquad \mathbf{B}_{12} = \begin{bmatrix} 3 & 8 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$$

$$\mathbf{B}_{21} = \begin{bmatrix} 9 & 1 & 4 & 6 \\ 6 & 8 & 1 & 4 \end{bmatrix}, \qquad \mathbf{B}_{22} = \begin{bmatrix} 8 & 7 \\ 3 & 2 \end{bmatrix}. \tag{4.2}$$

Using the matrices in (4.2), the matrix **B** of (4.1) can now be written as a matrix of matrices:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}. \tag{4.3}$$

This specification of **B** is called a *partitioning* of **B**, and the matrices $\mathbf{B}_{11}$, $\mathbf{B}_{12}, \mathbf{B}_{21}, \mathbf{B}_{22}$ are said to be *submatrices* of **B**; further, **B** of (4.3) is called a *partitioned* matrix.

Note that $\mathbf{B}_{11}$ and $\mathbf{B}_{21}$ have the same number of columns, as do $\mathbf{B}_{12}$ and $\mathbf{B}_{22}$. Likewise $\mathbf{B}_{11}$ and $\mathbf{B}_{12}$ have the same number of rows, as do $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$. This is the usual method of partitioning, as expressed in the general case for an $r \times c$ matrix:

$$\mathbf{A}_{r\times c} = \begin{bmatrix} \mathbf{K}_{p\times q} & \mathbf{L}_{p\times(c-q)} \\ \mathbf{M}_{(r-p)\times q} & \mathbf{N}_{(r-p)\times(c-q)} \end{bmatrix}$$

where $\mathbf{K}, \mathbf{L}, \mathbf{M}$, and $\mathbf{N}$ are the submatrices with their orders shown as subscripts.

Partitioning is not restricted to dividing a matrix into just four submatrices; it can be divided into numerous rows and columns of matrices. Thus if

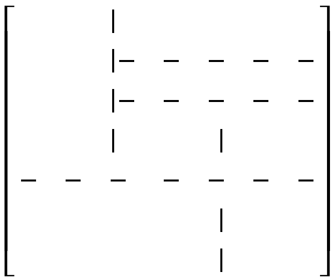
$$\mathbf{B}_{01} = \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \end{bmatrix}, \qquad \mathbf{B}_{02} = \begin{bmatrix} 3 & 8 \\ 1 & 1 \end{bmatrix},$$

$$\mathbf{B}_{03} = \begin{bmatrix} 4 & 3 & 6 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{B}_{04} = \begin{bmatrix} 2 & 1 \end{bmatrix}$$

with $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ as in (4.2), then $\mathbf{B}$ can be written in partitioned form as

$$\mathbf{B} = \left[\begin{array}{cccc|cc} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ \hline 4 & 3 & 6 & 1 & 2 & 1 \\ \hline 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{array}\right] = \begin{bmatrix} \mathbf{B}_{01} & \mathbf{B}_{02} \\ \mathbf{B}_{03} & \mathbf{B}_{04} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}. \qquad (4.4)$$

It goes without saying that each such line must always go the full length (or breadth) of the matrix. Partitioning in any staggered manner such as



is not allowed.

4.2.2 General Specification

The example illustrates the simplicity of the operation of partitioning a matrix and, as is evident from (4.1) and (4.4), there is no single way in which any matrix can be partitioned. In general, a matrix $\mathbf{A}$ of order $p \times q$ can be partitioned into r rows and c columns of submatrices as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2c} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{bmatrix}$$

where $\mathbf{A}_{ij}$ is the submatrix in the ith row and jth column of submatrices. If the ith row of submatrices has p_i rows of elements and the jth column of submatrices has q_j columns, then $\mathbf{A}_{ij}$ has order $p_i \times q_j$, where

$$\sum_{i=1}^{r} p_i = p \quad \text{and} \quad \sum_{j=1}^{c} q_j = q.$$

4.2.3 Transposing a Partitioned Matrix

The transpose of a partitioned matrix is the transposed matrix of transposed submatrices:

$$[\mathbf{X}\ \mathbf{Y}]' = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Y}' \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} & \mathbf{F} \end{bmatrix}' = \begin{bmatrix} \mathbf{A}' & \mathbf{D}' \\ \mathbf{B}' & \mathbf{E}' \\ \mathbf{C}' & \mathbf{F}' \end{bmatrix}.$$

The reader should use the example in (4.1) to verify these results and to be assured that in general the transpose of $[\mathbf{X}\ \mathbf{Y}]$ is neither $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ nor $[\mathbf{X}'\ \mathbf{Y}']$. For example,

$$\begin{bmatrix} 2 & | & 8 & 9 \\ 3 & | & 7 & 4 \end{bmatrix}' = \begin{bmatrix} 2 & & 3 \\ - & - & - \\ 8 & & 7 \\ 9 & & 4 \end{bmatrix}.$$

4.2.4 Partitioning Into Vectors

Suppose $\mathbf{a}_j$ is the jth column of $\mathbf{A}_{r\times c}$. Then

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \cdots \mathbf{a}_j \cdots \mathbf{a}_c \end{bmatrix} \tag{4.5}$$

is partitioned into its c columns. Similarly,

$$\mathbf{A} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_i' \\ \vdots \\ \alpha_r' \end{bmatrix} \tag{4.6}$$

is partitioned into its r rows α_i' for $i = 1, \ldots, r$.

4.3 THE TRACE OF A MATRIX

The sum of the diagonal elements of a square matrix is called the *trace* of the matrix, written $tr(\mathbf{A})$; that is, for $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, \ldots, n$

For example,

$$tr(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

$$tr\begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

When **A** is not square, the trace is not defined; that is, it does not exist.

The trace of a transposed matrix is the same as the trace of the matrix itself:

$$tr(\mathbf{A}') = tr(\mathbf{A}).$$

For example,

$$tr\begin{bmatrix} 1 & 8 & 4 \\ 7 & 3 & -2 \\ 6 & 9 & -8 \end{bmatrix} = tr\begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

Also, by treating a scalar as a 1×1 matrix we have

$$\text{tr(scalar)} = \text{scalar}; \quad \text{for example,} \quad \text{tr}(13) = 13.$$

4.4 ADDITION

We introduce the operation of addition by means of an example:

The numbers of lunches served one Saturday at a country club are shown in Table 4.1. Let us write the data of the table as a 2×3 matrix:

$$\mathbf{A} = \begin{bmatrix} 98 & 24 & 42 \\ 39 & 15 & 22 \end{bmatrix}.$$

Then, with the same frame of reference, the data for the Sunday lunches might be

$$\mathbf{B} = \begin{bmatrix} 55 & 19 & 44 \\ 43 & 53 & 38 \end{bmatrix}.$$

Hence over the weekend the total number of members served a beef lunch is the sum of the elements in the first row and first column of each matrix, $98 + 55 = 153$; and the total

TABLE 4.1 Number of Lunches

	Beef	Fish	Omelet
Member	98	24	42
Guest	39	15	22

number of guests served a fish lunch is $15 + 53 = 68$. In this way the matrix of all such sums is

$$\begin{bmatrix} 98+55 & 24+19 & 42+44 \\ 39+43 & 15+53 & 22+38 \end{bmatrix} = \begin{bmatrix} 153 & 43 & 86 \\ 82 & 68 & 60 \end{bmatrix},$$

which represents the numbers of lunches served over the weekend. This is the matrix sum **A** plus **B**; it is the matrix formed by adding the matrices **A** and **B**, element by element. Hence, if we write $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$ the matrix representing the sum of **A** and **B** is

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c;$$

that is, the sum of two matrices is the matrix of sums, element by element.

It is evident from this definition that matrix addition can take place only when the matrices involved are of the same order; that is, two matrices can be added only if they have the same number of rows and the same number of columns, in which case they are said to be *conformable for addition*.

Having defined addition, we can now consider the transpose of a sum and the trace of a sum:

(i) The transpose of a sum of matrices is the sum of the transposed matrices, that is,

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

(ii) The trace of a sum of matrices is the sum of the traces, that is,

$$tr(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$$

provided **A** and **B** are square, of the same order.

It is left to the reader to create examples illustrating these results. Formal proofs are as follows.

(i) If

$$\mathbf{A} + \mathbf{B} = \mathbf{C} = \{c_{ij}\} = \{a_{ij} + b_{ij}\},$$

then

$$(\mathbf{A} + \mathbf{B})' = \mathbf{C}' = \{c'_{ij}\} = \{c_{ji}\} = \{a_{ji} + b_{ji}\} = \{a_{ji}\} + \{b_{ji}\} = \mathbf{A}' + \mathbf{B}'.$$

(ii) $tr(\mathbf{A} + \mathbf{B}) = \sum_{i=1}^{n} (a_{ii} + b_{ii}) = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = tr(\mathbf{A}) + tr(\mathbf{B}).$

The difference between two matrices can be defined in a similar manner. For example, if $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{B} = \{b_{ij}\}$ are two matrices of the same order, then $\boldsymbol{A} - \boldsymbol{B} = \{a_{ij} - b_{ij}\}$, that is, it is the matrix of differences, element by element.

4.5 SCALAR MULTIPLICATION

We have just described matrix addition. A simple use of it shows that

$$\mathbf{A} + \mathbf{A} = \{a_{ij}\} + \{a_{ij}\} = \{2a_{ij}\} = 2\mathbf{A}.$$

Extending this to the case where λ is a positive integer, we have

$$\lambda\mathbf{A} = \mathbf{A} + \mathbf{A} + \mathbf{A} + \cdots + \mathbf{A},$$

there being λ $\mathbf{A}$'s in the sum on the right. Carrying out these matrix additions gives

$$\lambda\mathbf{A} = \{\lambda a_{ij}\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c.$$

This result, extended to λ being any scalar, is the definition of *scalar multiplication of a matrix*. Thus the matrix $\mathbf{A}$ multiplied by the scalar λ is the matrix $\mathbf{A}$ with every element multiplied by λ. For example

$$3\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 6 & -21 \\ 9 & 15 \end{bmatrix} \qquad \text{and} \quad \lambda\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2\lambda & -7\lambda \\ 3\lambda & 5\lambda \end{bmatrix}.$$

Although $\lambda\mathbf{A}$ is the customary notation, rather than $\mathbf{A}\lambda$, the latter is more convenient when λ is a fraction; for example, $\mathbf{A}/17$ rather than $(1/17)\mathbf{A}$ or $\mathbf{A}(1/17)$. Thus

$$\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} /17 = \begin{bmatrix} 2/17 & -7/17 \\ 3/17 & 5/17 \end{bmatrix}.$$

4.6 EQUALITY AND THE NULL MATRIX

Two matrices are said to be equal when they are identical element by element. Thus $\mathbf{A} = \mathbf{B}$ when $\{a_{ij}\} = \{b_{ij}\}$, meaning that $a_{ij} = b_{ij}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$. If

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 6 & -4 \\ 2 & 0 & 1 \end{bmatrix},$$

$\mathbf{A}$ is equal to $\mathbf{B}$ but not equal to $\mathbf{C}$. It is also apparent that equality of two matrices has no meaning unless they are of the same order.

Combining the ideas of subtraction and equality leads to the definition of zero in matrix algebra. For when $\mathbf{A} = \mathbf{B}$, then $a_{ij} = b_{ij}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$, and so

$$\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ij}\} = \{0\} = \mathbf{0}.$$

The matrix on the right is a matrix of zeros, that is, every element is zero. Such a matrix is called a *null matrix*; it is the zero of matrix algebra and is sometimes referred to as a *zero matrix*. It is, of course, not a unique zero because corresponding to every matrix there is a null matrix of the same order. For example, null matrices of order 2×4 and 3×3 are, respectively,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

4.7 MULTIPLICATION

The method of multiplying matrices is developed by first considering the product of two vectors and then the product of a matrix and a vector. Each of these products is introduced by means of an illustration and is then given a formal definition.

4.7.1 The Inner Product of Two Vectors

Illustration. Consider buying supplies of experimental rats, mice, and rabbits for laboratory courses in the chemistry, biochemistry, nutrition, and physiology departments of a university. Suppose the price per animal of rats, mice, and rabbits in the home town is \$3, \$1, and \$10, and that the chemistry department needs 50, 100, and 30 animals, respectively. The total cost to the chemistry department is, very simply,

$$3(50) + 1(100) + 10(30) = 550 \tag{4.7}$$

dollars. Suppose the prices are written as a row vector $\mathbf{a}'$ and the numbers of animals needed as a column vector $\mathbf{x}$:

$$\mathbf{a}' = \begin{bmatrix} 3 & 1 & 10 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix}.$$

Then the total cost of the animals needed, 550, is the sum of products of the elements of $\mathbf{a}'$ each multiplied by the corresponding element of $\mathbf{x}$. This is the definition of the product $\mathbf{a}'\mathbf{x}$. It is written as

$$\mathbf{a}'\mathbf{x} = \begin{bmatrix} 3 & 1 & 10 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 3(50) + 1(100) + 10(30) = 550. \tag{4.8}$$

This example illustrates the general procedure for obtaining $\mathbf{a}'\mathbf{x}$: multiply each element of the row vector $\mathbf{a}'$ by the corresponding element of the column vector $\mathbf{x}$ and add the products. The sum is $\mathbf{a}'\mathbf{x}$. Thus if

$$\mathbf{a}' = [a_1 \quad a_2 \cdots a_n] \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

their product $\mathbf{a}'\mathbf{x}$ is defined as

$$\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i.$$

It is called the *inner product* of the vectors $\mathbf{a}$ and $\mathbf{x}$. It exists only when $\mathbf{a}$ and $\mathbf{x}$ have the same order; when they are not of the same order the product $\mathbf{a}'\mathbf{x}$ is undefined.

4.7.2 A Matrix–Vector Product

Illustration. For the preceding illustration, suppose the animal prices in a neighboring town were \$2, \$2, and \$8, respectively. Let us represent them by the row vector [2 2 8]. Then purchasing the chemistry department's requirements in the neighboring town would cost

$$\begin{bmatrix} 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 2(50) + 2(100) + 8(30) = 540 \tag{4.9}$$

dollars, calculated just like (4.8).

Now put the two sets of prices as the rows of a matrix,

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}.$$

Then the products (4.8) and (4.9) can be represented simultaneously as a single product of the matrix $\mathbf{A}$ and the vector $\mathbf{x}$:

$$\mathbf{Ax} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = \begin{bmatrix} 3(50) + 1(100) + 10(30) \\ 2(50) + 2(100) + 8(30) \end{bmatrix} = \begin{bmatrix} 550 \\ 540 \end{bmatrix}. \tag{4.10}$$

The result is a vector, its elements being the inner products (4.8) and (4.9). In terms of the rows of $\mathbf{A}$, this means that the elements of the vector $\mathbf{Ax}$ are derived in exactly the same way as the product $\mathbf{a}'\mathbf{x}$ developed earlier, using the successive rows of $\mathbf{A}$ as the vector $\mathbf{a}'$. The result is the product $\mathbf{Ax}$; that is, $\mathbf{Ax}$ is obtained by repetitions of the product $\mathbf{a}'\mathbf{x}$ using the rows of $\mathbf{A}$ successively for $\mathbf{a}'$ and writing the results as a column vector. Hence, on using the notation of (4.6), with α_1' and α_2' being the rows of $\mathbf{A}$, we see that (4.10) is

$$\mathbf{Ax} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \end{bmatrix} \mathbf{x} = \begin{bmatrix} \alpha_1'\mathbf{x} \\ \alpha_2'\mathbf{x} \end{bmatrix}.$$

This generalizes at once to $\mathbf{A}$ having r rows:

$$\mathbf{Ax} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_i' \\ \vdots \\ \alpha_r' \end{bmatrix} \mathbf{x} = \begin{bmatrix} \alpha_1'\mathbf{x} \\ \alpha_2'\mathbf{x} \\ \vdots \\ \alpha_i'\mathbf{x} \\ \vdots \\ \alpha_r'\mathbf{x} \end{bmatrix}. \tag{4.11}$$

Thus $\mathbf{Ax}$ is a column vector, with its ith element being the inner product of the ith row of $\mathbf{A}$ with the column vector $\mathbf{x}$. Providing neither $\mathbf{A}$ nor $\mathbf{x}$ is a scalar, it is clear from this definition and from the example that $\mathbf{Ax}$ is defined only when the number of elements in each row of $\mathbf{A}$ (i.e., number of columns) is the same as the number of elements in the column vector $\mathbf{x}$, and when this occurs $\mathbf{Ax}$ is a column vector having the same number of elements as there are rows in $\mathbf{A}$. Therefore, when $\mathbf{A}$ has r rows and c columns and $\mathbf{x}$ is of order c, $\mathbf{Ax}$ is a column vector of order r; its ith element is $\Sigma_{k=1}^{c} a_{ik}x_k$ for $i = 1, 2, \ldots, r$. More formally, when

$$\mathbf{A} = \{a_{ij}\} \quad \text{and} \quad \mathbf{x} = \{x_j\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c,$$

then

$$\mathbf{Ax} = \left\{ \sum_{j=1}^{c} a_{ij}x_j \right\} \quad \text{for} \quad i = 1, 2, \ldots, r.$$

For example

$$\begin{bmatrix} 4 & 2 & 1 & 3 \\ 2 & 0 & -4 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4(1) + 2(0) + & 1(-1) + 3(3) \\ 2(1) + 0(0) + & -4(-1) + 7(3) \end{bmatrix} = \begin{bmatrix} 12 \\ 27 \end{bmatrix}.$$

Each element of $\mathbf{Ax}$ is obtained by moving along a row of $\mathbf{A}$ and down the column $\mathbf{x}$, multiplying each pair of corresponding elements and adding the products. This is always the procedure for calculating the product $\mathbf{Ax}$. It is also, as we shall see, the basis for calculating the product $\mathbf{AB}$ of two matrices.

Typical of many uses of the matrix–vector product is one that occurs in the study of inbreeding, where what is known as the generation matrix is used to relate the frequencies of mating types in one generation to those in another. Kempthorne(1957, p. 108), for example, gives the result earlier stated by Fisher (1949) that if after one generation of full-sib mating $\mathbf{f}^{(1)}$ represents the vector of frequencies of the seven distinct types of mating possible in this situation, then $\mathbf{f}^{(1)} = \mathbf{A}\mathbf{f}^{(0)}$, where $\mathbf{f}^{(0)}$ is the vector of initial frequencies and where $\mathbf{A}$ is the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{2}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16} & 0 \\ 0 & \frac{4}{16} & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{16} & \frac{4}{16} \\ 0 & \frac{2}{16} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{16} & 0 & \frac{2}{4} & 0 & \frac{4}{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{16} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{6}{16} & \frac{8}{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{16} \end{bmatrix}.$$

This simply means that if $f_i^{(1)}$ is the relative frequency of the ith type of mating after a generation of full-sib mating (that is, the ith element of the vector $\mathbf{f}^{(1)}$) and if $f_i^{(0)}$ is the corresponding initial frequency [ith element of $\mathbf{f}^{(0)}$], then, for example,

$$f_1^{(1)} = f_1^{(0)} + \tfrac{2}{16} f_2^{(0)} + \tfrac{1}{4} f_4^{(0)} + \tfrac{1}{16} f_6^{(0)};$$

and as another example

$$f_4^{(1)} = \tfrac{8}{16} f_2^{(0)} + \tfrac{2}{4} f_4^{(0)} + \tfrac{4}{16} f_6^{(0)}.$$

These and five similar equations are implied in the vector equation $\mathbf{f}^{(1)} = \mathbf{A}\mathbf{f}^{(0)}$. The matrix $\mathbf{A}$ which represents the relationships between the two sets of frequencies is known in this context as the generation matrix.

4.7.3 A Product of Two Matrices

Multiplying two matrices can be explained as a simple repetitive extension of multiplying a matrix by a vector.

Illustration. Continuing the illustration of buying laboratory animals, suppose the biochemistry department needed 60 rats, 80 mice, and 40 rabbits. The costs of these, obtained in the manner of (4.10), are shown in the product

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 60 \\ 80 \\ 40 \end{bmatrix} = \begin{bmatrix} 660 \\ 600 \end{bmatrix}. \tag{4.12}$$

Similarly, if the nutrition department needed 90, 30, and 20 animals, and the physiology department needed 30, 20, and 10, respectively, their costs would be

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}\begin{bmatrix} 90 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} 500 \\ 400 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}\begin{bmatrix} 30 \\ 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 210 \\ 180 \end{bmatrix}. \tag{4.13}$$

(The reader should verify the validity of these products.) Now, by writing the four column vectors of departmental requirements alongside one another as a matrix,

$$\mathbf{B} = \begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix},$$

the products in (4.10), (4.12), and (2.13) can be represented as the single matrix product

$$\mathbf{AB} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}\begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix} = \begin{bmatrix} 550 & 660 & 500 & 210 \\ 540 & 600 & 400 & 180 \end{bmatrix}. \tag{4.14}$$

This example illustrates how the product **AB** is simply a case of obtaining the product of **A** with each column of **B** and setting the products alongside one another. Thus for

$$\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \quad \mathbf{b}_4]$$

partitioned into columns in the manner of (4.5),

$$\mathbf{AB} = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \mathbf{Ab}_3 \quad \mathbf{Ab}_4].$$

Then, with **A** partitioned into its rows

$$\mathbf{A} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \end{bmatrix} \quad \text{we have} \quad \mathbf{AB} = \begin{bmatrix} \alpha_1'\mathbf{b}_1 & \alpha_1'\mathbf{b}_2 & \alpha_1'\mathbf{b}_3 & \alpha_1'\mathbf{b}_4 \\ \alpha_2'\mathbf{b}_1 & \alpha_2'\mathbf{b}_2 & \alpha_2'\mathbf{b}_3 & \alpha_2'\mathbf{b}_4 \end{bmatrix}.$$

Hence, the element of **AB** in its ith row and jth column is the inner product $\alpha_i'\mathbf{b}_j$ of the ith row of **A** and the jth column of **B**. And this is true in general:

$$\begin{aligned} \mathbf{A}_{r\times c}\mathbf{B}_{c\times s} &= \{\alpha_i'\mathbf{b}_j\} \\ &= \{a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ic}b_{cj}\} \\ &= \left\{\sum_{k=1}^{c} a_{ik}b_{kj}\right\}, \quad \text{for} \quad i = 1, \ldots, r \quad \text{and} \quad j = 1, \ldots, s. \end{aligned} \tag{4.15}$$

The (ij)th element of **AB** can therefore be obtained by thinking of moving from element to element along the ith row of **A** and simultaneously down the jth column of **B**, summing the products of corresponding elements. The resulting sum is the (ij)th element of **AB**. Schematically the operation can be represented as follows: to obtain **AB**,

$$\left[\overrightarrow{i\text{th row}}\right]_{r\times c}\left[\left\downarrow\begin{matrix} j\text{th}\\ \text{column}\end{matrix}\right.\right]_{c\times s} = \left[\left\{\begin{matrix}(ij)\text{th}\\ \text{element}\end{matrix}\right\}\right]_{r\times s}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$. The arrows indicate moving along the ith row and simultaneously down the jth column, summing the products of corresponding elements to get the (ij)th element of the product.

Once again, this is a matrix operation defined only if a certain condition is met: the ith row of **A** (and hence all rows) must have the same number of elements as does the jth column of **B** (and hence all columns). Since the number of elements in a row of a matrix is the number of columns in the matrix (and the number of elements in a column is the number of rows), this means that there must be exactly as many columns in **A** as there are rows in **B**. Thus the matrix product **AB** is defined only if the number of columns of **A** equals the number of rows of **B**. Note also, particularly in the numerical examples, that the product **AB** has the same number of rows as **A** and the same number of columns as **B**. This is true in general.

The important consequences of the definition of matrix multiplication are therefore as follows. The product **AB** of two matrices **A** and **B** is defined and therefore exists only if the number of columns in **A** equals the number of rows in **B**; the matrices are then said to be *conformable for multiplication for the product* **AB**, and **AB** has the same number of rows as **A** and the same number of columns as **B**. And the (ij) th element of **AB** is the inner product of the ith row of **A** and jth column of **B**.

Example 4.1 *For*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 4 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 0 & 6 & 1 & 5 \\ 1 & 1 & 0 & 7 \\ 3 & 4 & 4 & 3 \end{bmatrix}$$

the element in the first row and first column of the product ***AB*** *is the inner product of the first row of* ***A*** *and the first column of* ***B*** *and is*

$$1(0) + 0(1) + 2(3) = 6;$$

that in the first row and second column is

$$1(6) + 0(1) + 2(4) = 14;$$

and the element of ***AB*** *in the second row and third column is*

$$-1(1) + 4(0) + 3(4) = 11.$$

*In this way **AB** is obtained as*

$$\boldsymbol{AB} = \begin{bmatrix} 6 & 14 & 9 & 11 \\ 13 & 10 & 11 & 32 \end{bmatrix}.$$

The reader should verify this result.

4.7.4 Existence of Matrix Products

Using subscript notation the product, $\boldsymbol{P}$, of two matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, of orders $r \times c$ and $c \times s$, respectively, can be written as

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s},$$

a form which provides both for checking the conformability of **A** and **B** and for ascertaining the order of their product. Repeated use of this also simplifies determining the order of a matrix derived by multiplying several matrices together. Adjacent subscripts (which must be equal for conformability) "cancel out," leaving the first and last subscripts as the order of the product. For example, the product

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times s}\mathbf{C}_{s\times t}\mathbf{D}_{t\times u}$$

is a matrix of order $r \times u$.

This notation also demonstrates what is by now readily apparent from the definition of matrix multiplication, namely that the product **BA** does not necessarily exist, even if **AB** does. For **BA** can be written as $\mathbf{B}_{c\times s}\mathbf{A}_{r\times c}$, which we see at once is a legitimate product only if $s = r$. Otherwise **BA** is not defined. There are therefore three situations regarding the product of two matrices **A** and **B**. If **A** is of order $r \times c$:

(i) **AB** exists only if **B** has c rows.
(ii) **BA** exists only if **B** has r columns.
(iii) **AB** and **BA** both exist only if **B** is $c \times r$.

A corollary to (iii) is that $\mathbf{A}^2$ exists only when **A** is square. Another corollary is that both **AB** and **BA** always exist and are of the same order when **A** and **B** are square and of the same order. But as shall be shown subsequently, the two products are not necessarily equal. Their inequality will be discussed when considering the commutative law of multiplication, but meanwhile we simply state that they are not in general equal.

As a means of distinction, **AB** is described as **A** *postmultiplied* by **B**, or as **A** *multiplied on the right* by **B**; and **BA** is either **A** *premultiplied* by **B**, or **A** *multiplied on the left* by **B**.

4.7.5 Products With Vectors

Both the inner product of two vectors and the product of a matrix postmultiplied by a column vector are special cases of the general matrix product $\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s}$. Thus for the inner product of two vectors, $r = 1$ and $s = 1$, which means that $\mathbf{A}_{r\times c}$ becomes $\mathbf{A}_{1\times c}$, a row vector $(\mathbf{a}')_{1\times c}$, say; and $\mathbf{B}_{c\times s}$ becomes $\mathbf{B}_{c\times 1}$, a column vector $\mathbf{b}_{c\times 1}$. Thus we have the inner product

$$(\mathbf{a}')_{1\times c}\mathbf{b}_{c\times 1} = \mathbf{p}_{1\times 1},$$

a scalar. And the product in reverse order, where $\mathbf{a}$ and $\mathbf{b}$ can now be of different orders, is

$$\mathbf{b}_{c\times 1}(\mathbf{a}')_{1\times r} = \mathbf{P}_{c\times r},$$

a matrix: it is called the *outer product of* $\mathbf{b}$ *and* $\mathbf{a}'$. For

$$\mathbf{A}_{r\times c}\mathbf{b}_{c\times 1} = \mathbf{p}_{r\times 1},$$

the product being a column vector; similarly, a row vector postmultiplied by a matrix is a row vector:

$$(\mathbf{a}')_{1\times c}\mathbf{B}_{c\times r} = \mathbf{p}'_{1\times r}.$$

In words, these four results (for comformable products) are as follows:

(i) A row vector postmultiplied by a column vector is a scalar.
(ii) A column vector postmultiplied by a row vector is a matrix.
(iii) A matrix postmultiplied by a column vector is a column vector.
(iv) A row vector postmultiplied by a matrix is a row vector.

Example 4.2 *Given*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix}, \quad \boldsymbol{a}' = \begin{bmatrix} 1 & 5 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

then

$$\text{(i) } \boldsymbol{a}'\boldsymbol{b} = \begin{bmatrix} 1 & 5 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = 8.$$

$$\text{(ii) } \boldsymbol{b}\boldsymbol{a}' = \begin{bmatrix} 3 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 15 \\ 1 & 5 \end{bmatrix}.$$

$$\text{(iii) } \boldsymbol{A}\boldsymbol{b} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 25 \end{bmatrix}.$$

$$\text{(iv) } \boldsymbol{a}'\boldsymbol{B} = \begin{bmatrix} 1 & 5 \end{bmatrix}\begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1\ 13\ 26 \end{bmatrix}.$$

Illustration (Markov Chain). In the taxicab illustration of Example 2.1, suppose the driver lives in Town 1 and starts work from there each day. Denote this by a vector $\mathbf{x}'_0 = [1 \quad 0]$. In general $\mathbf{x}'$ is called a *state probability vector* (or simply *state vector*). The subscript 0 represents, the beginning of the day, and the elements 1 and 0 are probabilities of starting in Town 1 and in Town 2, respectively. This being so, the probabilities of being in Town

1 or Town 2 after the morning's first fare are, using the transition probability matrix **P** of equation (2.5).

$$\mathbf{x}_1' = \mathbf{x}_0'\mathbf{P} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 \end{bmatrix}. \tag{4.16}$$

And after the second fare the probabilities are

$$\mathbf{x}_2' = \mathbf{x}_1'\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.36 & 0.64 \end{bmatrix}. \tag{4.17}$$

Illustration (Linear Programming). An order for 1600 boxes of coronas, 500 boxes of half coronas, and 2000 boxes of cigarillos comes to a cigar-making company. The daily production and operating costs of the company's two small plants are shown in the following table. The company president needs to decide how many days each plant should operate to fill the order.

	Daily Production (100s of boxes)		
	Plant 1	Plant 2	Order Received
Product			
Corona	8	2	16
Half corona	1	1	5
Cigarillo	2	7	20
Daily operating cost	\$1000	\$2000	

The decision will be made on the criterion of filling the order with minimum cost. Let Plant 1 operate for x_1 days and Plant 2 for x_2 days to fill the order. Then to minimize cost we need to

$$\text{Minimize } f = 1000x_1 + 2000x_2,$$

and to fill the order x_1 and x_2 must satisfy

$$\begin{aligned} 8x_1 + 2x_2 &\geq 16 \\ x_1 + x_2 &\geq 5 \\ 2x_1 + 7x_2 &\geq 20, \end{aligned}$$

and because plants cannot operate for negative days we must also have

$$x_1 \geq 0 \quad \text{and} \quad x_2 \geq 0.$$

This is an example of what is called a linear programming problem. It can be stated generally as

$$\text{Minimize } f = \mathbf{c}'\mathbf{x}, \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \geq \mathbf{r} \quad \text{and} \quad \mathbf{x} \geq 0 \tag{4.18}$$

where the inequality sign is used in vector and matrix statements in exactly the same manner as the equality sign: $\mathbf{u} \geq \mathbf{v}$ means each element of $\mathbf{u}$ is equal to or greater than the corresponding element of $\mathbf{v}$, that is, $u_i \geq v_i$ for all i. In our example (4.18) has

$$\mathbf{c} = \begin{bmatrix} 1000 \\ 2000 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 8 & 2 \\ 1 & 1 \\ 2 & 7 \end{bmatrix}, \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 16 \\ 5 \\ 20 \end{bmatrix}.$$

Linear programming has its own vast literature with many problems involving more than just two variables (as does our example), solution of which can be quite difficult. Nevertheless, being able to state problems succinctly in matrix terminology, as in (4.18), provides a basis for being able to work efficiently with characteristics of a problem that lead to its solution. For example, it is known that (4.18) is equivalent to the problem

$$\text{Maximize } g = \mathbf{r}'\mathbf{z}, \quad \text{subject to} \quad \mathbf{A}'\mathbf{z} \leq \mathbf{c} \quad \text{and} \quad \mathbf{z} \geq \mathbf{0}. \tag{4.19}$$

Furthermore, if $\mathbf{x}_0$ and $\mathbf{z}_0$ are solutions of (4.18) and (4.19), respectively, then the minimum f in (4.18) and the maximum g in (4.19) are equal: $f_{\min} = \mathbf{c}'\mathbf{x}_0 = g_{\max} = \mathbf{r}'\mathbf{z}_0$. Derivation of such results and development of solutions $\mathbf{x}_0$ and $\mathbf{z}_0$ are beyond the scope of this book, but the ability to state complicated problems and relationships in terms of simple matrix products is to be noted.

4.7.6 Products With Scalars

To the extent that a scalar can be considered a 1×1 matrix, certain cases of the scalar multiplication of a matrix in Section 4.4 are included in the general matrix product $\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s}$, to give $a_{1\times 1}\mathbf{b}'_{1\times c} = a\mathbf{b}'$ and $(\mathbf{a}')_{r\times 1}b_{1\times 1} = b\mathbf{a}'$.

4.7.7 Products With Null Matrices

For any matrix $\mathbf{A}_{r\times s}$, pre- or post-multiplication by a null matrix of appropriate order results in a null matrix. Thus if $\mathbf{0}_{c\times r}$ is a null matrix of order $c \times r$,

$$\mathbf{0}_{c\times r}\mathbf{A}_{r\times s} = \mathbf{0}_{c\times s} \quad \text{and} \quad \mathbf{A}_{r\times s}\mathbf{0}_{s\times p} = \mathbf{0}_{r\times p}.$$

For example,

$$\begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

In writing this equality as $\mathbf{0A} = \mathbf{0}$, it is important to notice that the two $\mathbf{0}$'s are not of the same order.

4.7.8 Products With Diagonal Matrices

A diagonal matrix is defined as a square matrix having all off-diagonal elements zero. Multiplication by a diagonal matrix is particularly easy: premultiplication of a matrix $\mathbf{A}$ by a diagonal matrix $\mathbf{D}$ gives a matrix whose rows are those of $\mathbf{A}$ multiplied by the respective diagonal elements of $\mathbf{D}$. For example, in the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}$$

of (4.10), rows represent prices of rats, mice, and rabbits bought locally or in a neighboring town. Suppose prices increase 5% locally and 20% in the neighboring town. Then with

$$\mathbf{D} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix}$$

the matrix of new prices is

$$\mathbf{DA} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix} \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 3.15 & 1.05 & 10.50 \\ 2.40 & 2.40 & 9.60 \end{bmatrix}.$$

4.7.9 Identity Matrices

A diagonal matrix having all diagonal elements equal to unity is called an *identity matrix*, or sometimes a *unit matrix*. It is usually denoted by the letter $\mathbf{I}$, with a subscript for its order when necessary for clarity; for example,

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When $\mathbf{A}$ is of order $p \times q$

$$\mathbf{I}_p\mathbf{A}_{p\times q} = \mathbf{A}_{p\times q}\mathbf{I}_q = \mathbf{A}_{p\times q};$$

that is, multiplication of a matrix $\mathbf{A}$ by an (conformable) identity matrix does not alter $\mathbf{A}$. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix}.$$

4.7.10 The Transpose of a Product

The transpose of a product matrix is the product of the transposed matrices taken in reverse sequence, that is, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Example 4.3

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 3 & 0 & 7 \end{bmatrix} = \begin{bmatrix} -2 & 1 & -6 \\ 11 & 0 & 19 \end{bmatrix}$$

$$\boldsymbol{B}'\boldsymbol{A}' = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 2 & 0 \\ 1 & 4 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 11 \\ 1 & 0 \\ -6 & 9 \end{bmatrix} = (\boldsymbol{AB})'.$$

Consideration of order and conformability for multiplication confirms this result. If $\mathbf{A}$ is $r \times s$ and $\mathbf{B}$ is $s \times t$, the product $\mathbf{P} = \mathbf{AB}$ is $r \times t$; that is, $\mathbf{A}_{r\times s}\mathbf{B}_{s\times t} = \mathbf{P}_{r\times t}$. But $\mathbf{A}'$ is $s \times r$ and $\mathbf{B}'$ is $t \times s$ and the only product to be derived from these is $(\mathbf{B}')_{t\times s}(\mathbf{A}')_{s\times r} = \mathbf{Q}_{t\times r}$ say. That $\mathbf{Q} = \mathbf{B}'\mathbf{A}'$ is indeed the transpose of $\mathbf{P} = \mathbf{AB}$ is apparent from the definition of multiplication: the (ij)th term of $\mathbf{Q}$ is the inner product of the ith row of $\mathbf{B}'$ and the jth column of $\mathbf{A}'$, which in turn is the inner product of the ith column of $\mathbf{B}$ and the jth row of $\mathbf{A}$, and this by the definition of multiplication is the (ji)th term of $\mathbf{P}$. Hence, $\mathbf{Q} = \mathbf{P}'$, or $\mathbf{B}'\mathbf{A}' = (\mathbf{AB})'$. More formally,

$$\mathbf{AB} = \mathbf{P} = \{p_{ij}\} = \left\{ \sum_{k=1}^{s} a_{ik}b_{kj} \right\},$$

This result for the transpose of the product of two matrices extends directly to the product of more than two. For example, $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$ and $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$. Proof is left as an exercise for the reader.

4.7.11 The Trace of a Product

The trace of a matrix is defined in Section 4.3. Thus if $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, the trace of $\mathbf{A}$ is $tr(\mathbf{A}) = \Sigma_{i=1}^{n} a_{ii}$, the sum of the diagonal elements. We now show that for the product $\mathbf{AB}$, $tr(\mathbf{AB}) = tr(\mathbf{BA})$ and hence $tr(\mathbf{ABC}) = tr(\mathbf{BCA}) = tr(\mathbf{CAB})$. Note that $tr(\mathbf{AB})$ exists only if $\mathbf{AB}$ is square, which occurs only when $\mathbf{A}$ is $r \times c$ and $\mathbf{B}$ is $c \times r$. Then if $\mathbf{AB} = \mathbf{P} = \{p_{ij}\}$ and $\mathbf{BA} = \mathbf{T} = \{t_{ij}\}$,

$$\begin{aligned} tr(\mathbf{AB}) &= \sum_{i=1}^{r} p_{ii} = \sum_{i=1}^{r} \left(\sum_{j=1}^{c} a_{ij}b_{ji} \right) = \sum_{i=1}^{r} \left(\sum_{j=1}^{c} b_{ji}a_{ij} \right) \\ &= \sum_{j=1}^{c} \left(\sum_{i=1}^{r} b_{ji}a_{ij} \right) = \sum_{j=1}^{c} t_{jj} = \mathrm{tr}(\mathbf{BA}). \end{aligned}$$

Extension to products of three or more matrices is obvious.

Notice that the intermediate result

$$tr(\mathbf{AB}) = \sum_{i=1}^{r} \sum_{j=1}^{c} a_{ij}b_{ji}$$

can also, using $\mathbf{B}' = \{b'_{ij} = b_{ji}\}$, be expressed as

$$tr(\mathbf{AB}) = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}b'_{ij},$$

which is the sum of products of each element of $\mathbf{A}$ multiplied by the corresponding element of $\mathbf{B}'$. And if $\mathbf{B} = \mathbf{A}'$, we have

$$tr(\mathbf{AA}') = tr(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}^2,$$

that the trace of $\mathbf{AA}'$ (and of $\mathbf{A}'\mathbf{A}$) is the sum of squares of elements of $\mathbf{A}$. It is left as an exercise for the reader to verify these results numerically.

4.7.12 Powers of a Matrix

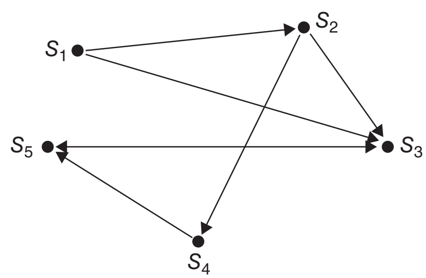
Since $\mathbf{A}_{r\times c}\mathbf{A}_{r\times c}$ exists only if $r = c$, that is, only if $\mathbf{A}$ is square, we see that $\mathbf{A}^2$ exists only when $\mathbf{A}$ is square; and then $\mathbf{A}^k$ exists for all positive integers k. And, in keeping with scalar arithmetic where $x^0 = 1$, we take $\mathbf{A}^0 = \mathbf{I}$ for $\mathbf{A}$ square.

For example, in the taxicab illustration, it is clear from (4.16) that (4.17) is

$$\begin{aligned}\mathbf{x}'_2 = \mathbf{x}'_1\mathbf{P} = \mathbf{x}'_0\mathbf{P}^2 &= \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0.2 & 0.8\\ 0.4 & 0.6\end{bmatrix}^2 \\ &= \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0.2 & 0.8\\ 0.4 & 0.6\end{bmatrix}\begin{bmatrix}0.2 & 0.8\\ 0.4 & 0.6\end{bmatrix} = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0.36 & 0.64\\ 0.32 & 0.68\end{bmatrix} = \begin{bmatrix}0.36 & 0.64\end{bmatrix}.\end{aligned}$$

Similarly, $\mathbf{x}'_3 = \mathbf{x}'_0\mathbf{P}^3$, and in general $\mathbf{x}'_n = \mathbf{x}'_0\mathbf{P}^n$.

Illustration (Graph Theory). Suppose in a communications network of five stations messages can be sent only in the directions of the arrows of the accompanying diagram.



This chart of the possible message routes can also be represented by a matrix $\mathbf{T} = \{t_{ij}\}$ say, where $t_{ij} = 0$ except $t_{ij} = 1$ if a message can be sent from S_i to S_j. Hence,

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

In the rth power of $\mathbf{T}$, say $\mathbf{T}^r = \{t_{ij}^{(r)}\}$, the element $t_{ij}^{(r)}$ is then the number of ways of getting a message from station i to station j in exactly r steps. Thus,

$$\mathbf{T}^2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{T}^3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

show that messages can be transmitted in two ways from S_2 to S_5 in two steps and in two ways from S_2 to S_3 in three steps. In this manner the pathways through a network can be counted simply by looking at powers of $\mathbf{T}$. More than that, if there is no direct path from S_i to S_j, that is, $t_{ij} = 0$, the powers of $\mathbf{T}$ can be used to ascertain if there are any indirect paths. For example, $t_{25} = 0$ but $t_{25}^{(2)} = 2$, showing that the path from S_2 to S_5 cannot be traversed directly but there are two indirect routes of two steps each. Thus only if $\Sigma_{r=0}^{\infty} t_{ij}^{(r)} = 0$ is there no path at all from S_i to S_j.

Applications of matrices like $\mathbf{T}$ arise in many varied circumstances. For example, instead of the S's being stations in a communications network, they could be people in a social or business group, whereupon each arrow of the figure could represent the dominance of one person by another. Then $\mathbf{T}^2$ represents the two-stage dominance existing within the group of people; and so on. Or the figure could represent the spreading of a rumor between people or groups of people; or it could be the winning of sports contests; and so on. In this way, graph theory, supported by matrix representations and attendant matrix algebra, can provide insight to a host of real-life problems.

4.7.13 Partitioned Matrices

When matrices $\mathbf{A}$ and $\mathbf{B}$ are partitioned so that their submatrices are appropriately conformable for multiplication, the product $\mathbf{AB}$ can be expressed in partitioned form having submatrices that are functions of the submatrices of $\mathbf{A}$ and $\mathbf{B}$. For example, if

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{bmatrix},$$

then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \end{bmatrix}.$$

This means that the partitioning of **A** along its columns must be the same as that of **B** along its rows. Then $\mathbf{A}_{11}$ (and $\mathbf{A}_{21}$) has the same number of columns as $\mathbf{B}_{11}$ has rows, and $\mathbf{A}_{12}$ (and $\mathbf{A}_{22}$) has the same number of columns as $\mathbf{B}_{21}$ has rows. We see at once that when two matrices are appropriately partitioned the submatrices of their product are obtained by treating the submatrices of each of them as elements in a normal matrix product, and the individual elements of the product are derived in the usual way from the products of the submatrices.

In general, if **A** is $p \times q$ and is partitioned as

$$\mathbf{A}_{p\times q} = \{\mathbf{A}_{ij}(p_i \times q_j)\}, \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c$$

with $\Sigma_{i=1}^{r} p_\iota = p$ and $\Sigma_{J=1}^{c} q_j = q$, where $p_\iota \times q_j$ is the order of the submatrix $\mathbf{A}_{\iota j}$, and likewise if

$$\mathbf{B}_{q\times s} = \{\mathbf{B}_{jk}(q_J \times s_k)\}, \quad \text{for} \quad j = 1, 2, \ldots, c \quad \text{and} \quad k = 1, 2, \ldots, d$$

with $\Sigma_{J=1}^{c} q_J = q$ and $\Sigma_{k=1}^{d} s_k = s$, then

$$(\mathbf{AB})_{p\times s} = \left\{ \sum_{j=1}^{c} \mathbf{A}_{iJ}\mathbf{B}_{jk}(p_\iota \times s_k) \right\},$$

for $i = 1, 2, \ldots, r$ and $k = 1, 2, \ldots, d$.

Illustration. Feller (1968, p. 439) gives an example of a matrix of transition probabilities that can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{U}_1 & \mathbf{V}_1 & \mathbf{T} \end{bmatrix},$$

where **A** and **B** are also transition probability matrices. It is readily shown that for $\mathbf{U}_n = \mathbf{U}_1\mathbf{A}^{n-1} + \mathbf{T}\mathbf{U}_{n-1}$ and $\mathbf{V}_n = \mathbf{V}_1\mathbf{B}^{n-1} + \mathbf{T}\mathbf{V}_{n-1}$,

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{A}^n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^n & \mathbf{0} \\ \mathbf{U}_n & \mathbf{V}_n & \mathbf{T}^n \end{bmatrix}.$$

4.7.14 Hadamard Products

The definition of a product that has been presented at some length is the one most generally used. But because matrices are arrays of numbers, they provide opportunity for defining products in several different ways.

The Hadamard product of matrices $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ is defined only when $\mathbf{A}$ and $\mathbf{B}$ have the same order. It is the matrix of the element-by-element products of corresponding elements in $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A} \cdot \mathbf{B} = \{a_{ij}b_{ij}\}.$$

Thus the (ij)th element of the Hadamard product $\mathbf{A} \cdot \mathbf{B}$ is the product of the (ij)th elements of $\mathbf{A}$ and $\mathbf{B}$:

$$\begin{bmatrix} 2 & 0 & -1 \\ 4 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 7 \\ 2 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -7 \\ 8 & 0 & 18 \end{bmatrix}.$$

For example, suppose that in the taxicab illustration that the average fare between the two towns is \$8 and within Town 1 it is \$6 and within Town 2 it is \$5. Then, corresponding to the transition probability matrix, the reward matrix is

$$\mathbf{R} = \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix}.$$

Therefore, since the expected reward for a fare from Town i to Town j is $p_{ij}r_{ij}$, the matrix of expected rewards is the Hadamard product of $\mathbf{P}$ and $\mathbf{R}$:

$$\mathbf{P} \cdot \mathbf{R} = \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \cdot \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix} = \begin{bmatrix} 1.2 & 6.4 \\ 3.2 & 3.0 \end{bmatrix}. \tag{4.20}$$

4.8 THE LAWS OF ALGEBRA

We now give formal consideration to the associative, commutative, and distributive laws of algebra as they relate to the addition and multiplication of matrices.

4.8.1 Associative Laws

The addition of matrices is associative provided the matrices are conformable for addition. For if $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ have the same order,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \{a_{ij} + b_{ij}\} + \{c_{ij}\} = \{a_{ij} + b_{ij} + c_{ij}\} = \mathbf{A} + \mathbf{B} + \mathbf{C}.$$

Also,

$$\{a_{ij} + b_{ij} + c_{ij}\} = \{a_{ij}\} + \{b_{ij} + c_{ij}\} = \mathbf{A} + (\mathbf{B} + \mathbf{C}),$$

so proving the associative law of addition.

In general, the laws of algebra that hold for matrices do so because matrix results follow directly from corresponding scalar results for their elements—as illustrated here. Further proofs in this section are therefore omitted.

The associative law is also true for multiplication, provided the matrices are conformable for multiplication. For if **A** is $p \times q$, **B** is $q \times r$, and **C** is $r \times s$, then $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$.

4.8.2 The Distributive Law

The distributive law holds true. For example,

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

provided both **B** and **C** are conformable for addition (necessarily of the same order) and **A** and **B** are conformable for multiplication (and hence **A** and **C** also).

4.8.3 Commutative Laws

Addition of matrices is commutative (provided the matrices are conformable for addition). If **A** and **B** are of the same order

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} = \{b_{ij} + a_{ij}\} = \mathbf{B} + \mathbf{A}.$$

Multiplication of matrices is not in general commutative. As seen earlier, there are two possible products that can be derived from matrices **A** and **B**, **AB** and **BA**, and if **A** is of order $r \times c$ both products exist only if **B** is of order $c \times r$. **AB** is then square, of order $r \times r$, and **BA** is also square, of order $c \times c$. Possible equality of **AB** and **BA** can therefore be considered only where $r = c$, in which case **A** and **B** are both square and have the same order $r \times r$. The products are then

$$\mathbf{AB} = \left\{ \sum_{k=1}^{r} a_{ik} b_{kj} \right\} \quad \text{and} \quad \mathbf{BA} = \left\{ \sum_{k=1}^{r} b_{ik} a_{kj} \right\} \quad \text{for} \quad i, j = 1, \ldots, r.$$

It can be seen that the (ij)th elements of these products do not necessarily have even a single term in common in their sums of products, let alone are they equal. Therefore, even when **AB** and **BA** both exist and are of the same order, they are not in general equal; for example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 4 & -7 \end{bmatrix}$$
$$\neq \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ -2 & -2 \end{bmatrix}.$$

But in certain cases **AB** and **BA** *are* equal; for example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 12 & 18 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Two special cases of matrix multiplication being commutative are $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$ and $\mathbf{0A} = \mathbf{A0} = \mathbf{0}$ for $\mathbf{A}$ being square. When $\mathbf{A}$ is rectangular the former holds, with identity matrices of different orders, that is, $\mathbf{I}_r\mathbf{A}_{r\times c} = \mathbf{A}_{r\times c}\mathbf{I}_c = \mathbf{A}_{r\times c}$; the latter can be expressed more generally as two separate statements: $\mathbf{0}_{p\times r}\mathbf{A}_{r\times c} = \mathbf{0}_{p\times c}$ and $\mathbf{A}_{r\times c}\mathbf{0}_{c\times s} = \mathbf{0}_{r\times s}$.

4.9 CONTRASTS WITH SCALAR ALGEBRA

The definition of matrix multiplication leads to results in matrix algebra that have no counterpart in scalar algebra. In fact, some matrix results contradict their scalar analogues. We give some examples.

First, although in scalar algebra $ax + bx$ can be factored either as $x(a + b)$ or as $(a + b)x$, this duality is not generally possible with matrices:

$$\mathbf{AX} + \mathbf{BX} = (\mathbf{A} + \mathbf{B})\mathbf{X} \quad \text{and} \quad \mathbf{XA} + \mathbf{XB} = \mathbf{X}(\mathbf{A} + \mathbf{B}),$$

but

$$\mathbf{XP} + \mathbf{QX} \text{ generally does not have } \mathbf{X} \text{ as a factor.}$$

Another example of factoring is that, similar to $xy - x = x(y - 1)$ in scalar algebra, $\mathbf{XY} - \mathbf{X} = \mathbf{X}(\mathbf{Y} - \mathbf{I})$ in matrix algebra, but with the second term inside the parentheses being $\mathbf{I}$ and not the scalar 1. (It cannot be 1 because 1 is not conformable for subtraction from $\mathbf{Y}$.) Furthermore, the fact that it is $\mathbf{I}$ emphasizes, because of conformability for subtraction from $\mathbf{Y}$, that $\mathbf{Y}$ must be square. Of course this can also be gleaned directly from $\mathbf{XY} - \mathbf{X}$ itself: if $\mathbf{X}$ is $r \times c$, then $\mathbf{Y}$ must have c rows in order for $\mathbf{XY}$ to be defined, and $\mathbf{Y}$ must have c columns in order for $\mathbf{XY}$ and $\mathbf{X}$ to be conformable for subtraction. This simple example illustrates the need for constantly keeping conformability in mind, especially when matrix symbols do not have their orders attached.

Another consequence of matrix multiplication is that even when $\mathbf{AB}$ and $\mathbf{BA}$ both exist, they may not be equal. Thus for

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \tag{4.21}$$

we have

$$\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \mathbf{BA} = \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix}.$$

Notice here that even though $\mathbf{AB} = \mathbf{0}$, neither $\mathbf{A}$ nor $\mathbf{B}$ is $\mathbf{0}$. This illustrates an extremely important feature of matrices: the equation $\mathbf{AB} = \mathbf{0}$ does *not* always lead to the conclusion that $\mathbf{A}$ or $\mathbf{B}$ is $\mathbf{0}$, as would be the case with scalars. A further illustration arises from observing that for $\mathbf{A}$ and $\mathbf{B}$ of (4.21), we have $\mathbf{BA} = 2\mathbf{B}$. This can be rewritten as $\mathbf{BA} - 2\mathbf{B} = \mathbf{0}$, equivalent to $\mathbf{B}(\mathbf{A} - 2\mathbf{I}) = \mathbf{0}$. But from this we cannot conclude either that $\mathbf{A} - 2\mathbf{I}$ is $\mathbf{0}$ or that $\mathbf{B} = \mathbf{0}$. Similarly, not even the equation $\mathbf{X}^2 = \mathbf{0}$ means that $\mathbf{X} = \mathbf{0}$; for example, with

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \quad \text{we have} \quad \mathbf{X}^2 = \mathbf{0}.$$

Likewise $\mathbf{Y}^2 = \mathbf{I}$ implies neither $\mathbf{Y} = \mathbf{I}$ nor $\mathbf{Y} = -\mathbf{I}$; for example,

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 \\ 4 & -1 \end{bmatrix}, \quad \text{but} \quad \mathbf{Y}^2 = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly, we can have $\mathbf{M}^2 = \mathbf{M}$ with both $\mathbf{M} \neq \mathbf{I}$ and $\mathbf{M} \neq \mathbf{0}$; for example,

$$\mathbf{M} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} = \mathbf{M}^2.$$

Such a matrix $\mathbf{M}$ with $\mathbf{M} = \mathbf{M}^2$ is said to be *idempotent*.

4.10 DIRECT SUM OF MATRICES

Direct sums and direct products are matrix operations defined in terms of partitioned matrices. They are discussed in this and the next section.

The direct sum of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \tag{4.22}$$

and extends very simply to more than two matrices:

$$\mathbf{A} \oplus \mathbf{B} \oplus \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix}$$

and

$$\bigoplus_{i=1}^{k} \mathbf{A}_i = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & & \vdots \\ \vdots & & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_k \end{bmatrix} = \operatorname{diag}\{\mathbf{A}_i\} \text{ for } i = 1, \ldots, k.$$

The definition (4.22) and its extensions apply whether or not the submatrices are of the same order; and all null matrices are of appropriate order. For example,

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \oplus \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 8 & 9 \end{bmatrix}. \tag{4.23}$$

Transposing a direct sum gives the direct sum of the transposes. It is clear from (4.22) that $\mathbf{A} \oplus (-\mathbf{A}) \neq \mathbf{0}$ unless $\mathbf{A}$ is null. Also,

$$(\mathbf{A} \oplus \mathbf{B}) + (\mathbf{C} \oplus \mathbf{D}) = (\mathbf{A} + \mathbf{C}) \oplus (\mathbf{B} + \mathbf{D})$$

only if the necessary conditions of conformability for addition are met. Similarly,

$$(\mathbf{A} \oplus \mathbf{B})(\mathbf{C} \oplus \mathbf{D}) = \mathbf{AC} \oplus \mathbf{BD}$$

provided that conformability for multiplication is satisfied. The direct sum $\mathbf{A} \oplus \mathbf{B}$ is square only if $\mathbf{A}$ is $p \times q$ and $\mathbf{B}$ is $q \times p$. The determinant of $\mathbf{A} \oplus \mathbf{B}$ equals $|\mathbf{A}||\mathbf{B}|$ if both $\mathbf{A}$ and $\mathbf{B}$ are square, but otherwise it is zero or nonexistent.

4.11 DIRECT PRODUCT OF MATRICES

The direct product of two matrices $\mathbf{A}_{p \times q}$ and $\mathbf{B}_{m \times n}$ is defined as

$$\mathbf{A}_{p \times q} \otimes \mathbf{B}_{m \times n} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix} \tag{4.24}$$

and is sometimes called the Kronecker product or Zehfuss product [(see Henderson et al. (1981)]. Clearly, (4.24) is partitioned into as many submatrices as there are elements of $\mathbf{A}$, each submatrix being $\mathbf{B}$ multiplied by an element of $\mathbf{A}$. Therefore the elements of the direct product consist of all possible products of an element of $\mathbf{A}$ multiplied by an element of $\mathbf{B}$. It has order $pm \times qn$. For example,

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \otimes \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 7 & 12 & 14 & 18 & 21 \\ 8 & 9 & 16 & 18 & 24 & 27 \end{bmatrix}.$$

The transpose of a direct product is the direct product of the transposes—as is evident from transposing (4.24).

Direct products have many useful and interesting properties, some of which are as follows.

(i) In contrast to $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, we have $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.

(ii) For $\mathbf{x}$ and $\mathbf{y}$ being vectors: $\mathbf{x}' \otimes \mathbf{y} = \mathbf{yx}' = \mathbf{y} \otimes \mathbf{x}'$.

(iii) For λ being a scalar: $\lambda \otimes \mathbf{A} = \lambda\mathbf{A} = \mathbf{A} \otimes \lambda = \mathbf{A}\lambda$.

(iv) For partitioned matrices, although

$$[\mathbf{A}_1 \quad \mathbf{A}_2] \otimes \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B} \quad \mathbf{A}_2 \otimes \mathbf{B}],$$

$$\mathbf{A} \otimes [\mathbf{B}_1 \quad \mathbf{B}_2] \neq [\mathbf{A} \otimes \mathbf{B}_1 \quad \mathbf{A} \otimes \mathbf{B}_2].$$

(v) Provided conformability requirements for regular matrix multiplication are satisfied, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{X} \otimes \mathbf{Y}) = \mathbf{AX} \otimes \mathbf{BY}$.

(vi) For $\mathbf{D}_k$ being a diagonal matrix of order $k \times k$ with elements d_i: $\mathbf{D}_k \otimes \mathbf{A} = d_1\mathbf{A} \oplus d_2\mathbf{A} \oplus \cdots \oplus d_k\mathbf{A}$.

(vii) The trace obeys product rules: $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

(viii) Provided $\mathbf{A}$ and $\mathbf{B}$ are square, $|\mathbf{A}_{p\times p} \otimes \mathbf{B}_{m\times m}| = |\mathbf{A}|^m|\mathbf{B}|^p$.

These results are readily illustrated with simple numerical examples, and most of them are not difficult to prove.

Sometimes $\mathbf{A} \otimes \mathbf{B}$ is referred to as the *right direct product* to distinguish it from $\mathbf{B} \otimes \mathbf{A}$ which is then called the *left direct product;* and on rare occasions the right-hand side of (4.24) will be found defined as $\mathbf{B} \otimes \mathbf{A}$. Whatever names are used, it is apparent from (4.24) that for $\mathbf{A}_{p\times q} = \{a_{ij}\}$ and $\mathbf{B}_{m\times n} = \{b_{rs}\}$ the elements of both $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{B} \otimes \mathbf{A}$ consist of all possible products $a_{ij}b_{rs}$. In fact, $\mathbf{B} \otimes \mathbf{A}$ is simply $\mathbf{A} \otimes \mathbf{B}$ with the rows and columns each in a different order.

Illustration. A problem in genetics that uses direct products is concerned with n generations of random mating starting with the progeny obtained from crossing two autotetraploid plants which both have genotype AAaa. Normally the original plants would produce gametes AA, Aa, and aa in the proportion 1: 4: 1. But suppose the proportion is $u : 1 - 2u : u$ where, for example, u might take the value $(1 - \alpha)/6$, for α being a measure of "diploidization" of the plants: $\alpha = 0$ is the case of autotetraploids with chromosome segregation and $\alpha = 1$ is the diploid case with all gametes being Aa. The question now is, what are the genotypic frequencies in the population after n generations of random mating? Let $\mathbf{u}_i$ be the vector of gametic frequencies and $\mathbf{f}_i$ the vector of genotype frequencies in the ith generation of random mating, where $\mathbf{u}_0$ is the vector of gametic frequencies in the initial plants. Then

$$\mathbf{u}_0 = \begin{bmatrix} u \\ 1 - 2u \\ u \end{bmatrix}$$

and $\mathbf{f}_{i+1} = \mathbf{u}_i \otimes \mathbf{u}_i$ for $i = 0, 1, 2, \ldots, n$. Furthermore, the relationship between $\mathbf{u}_i$ and $\mathbf{f}_i$ at any generation is $\mathbf{u}_i = \mathbf{Bf}_i$ where

$$\mathbf{B} = \begin{bmatrix} 1 & \frac{1}{2} & u & \frac{1}{2} & u & 0 & u & 0 & 0 \\ 0 & \frac{1}{2} & 1-2u & \frac{1}{2} & 1-2u & \frac{1}{2} & 1-2u & \frac{1}{2} & 0 \\ 0 & 0 & u & 0 & u & \frac{1}{2} & u & \frac{1}{2} & 1 \end{bmatrix}.$$

Thus

$$\begin{aligned} \mathbf{f}_i &= \mathbf{u}_{i-1} \otimes \mathbf{u}_{i-1} = \mathbf{Bf}_{i-1} \otimes \mathbf{Bf}_{i-1} = (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-1} \otimes \mathbf{f}_{i-1}) \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})] \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B}) \otimes (\mathbf{B} \otimes \mathbf{B})][(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})]. \end{aligned}$$

It is easily seen (and can be verified by induction) that

$$\mathbf{f}_i = \otimes^2\mathbf{B}(\otimes^4\mathbf{B})(\otimes^8\mathbf{B}) \cdots (\otimes^{2^{l-1}}\mathbf{B})(\otimes^{2^l}\mathbf{u}_0)$$

where $\otimes^n\mathbf{B}$ means the direct product of $n\mathbf{B}$'s.

4.12 THE INVERSE OF A MATRIX

Let $\boldsymbol{A} = (a_{ij})$ be a matrix of order $n \times n$ whose determinant is not equal to zero. This matrix is said to be *nonsingular*. The *inverse* of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$, is an $n \times n$ matrix that satisfies the condition $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$. Such a matrix is unique.

The inverse of $\boldsymbol{A}$ can be computed as follows: let c_{ij} denote the cofactor of a_{ij} (see Section 3.3.3). Define the matrix $\boldsymbol{C}$ as $\boldsymbol{C} = (c_{ij})$. The inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \frac{\boldsymbol{C}'}{|\boldsymbol{A}|},$$

where $\boldsymbol{C}'$ is the transpose of $\boldsymbol{C}$. The matrix $\boldsymbol{C}'$ is called the *adjugate*, or *adjoint*, matrix of $\boldsymbol{A}$, and is denoted by *adj* $\boldsymbol{A}$. For example, if $\boldsymbol{A}$ is the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 \\ 3 & 9 \end{bmatrix},$$

then

$$adj\,\boldsymbol{A} = \begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix},$$

and

$$\boldsymbol{A}^{-1} = \frac{1}{3}\begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix},$$

since the determinant of $\boldsymbol{A}$ is equal to 3.

The following are some properties associated with the inverse operation [see, e.g., Harville (1997, Chapter 8)]:

(a) $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
(b) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.
(c) $|\boldsymbol{A}^{-1}| = \frac{1}{|\boldsymbol{A}|}$, if $\boldsymbol{A}$ is nonsingular.
(d) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.
(e) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.
(f) $(\boldsymbol{A} \oplus \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \oplus \boldsymbol{B}^{-1}$.

(g) If $\boldsymbol{A}$ is partitioned as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},$$

where $\boldsymbol{A}_{ij}$ is of order $n_i \times n_j$ $(i,j = 1,2)$, then

$$|\boldsymbol{A}| = \begin{cases} |\boldsymbol{A}_{11}||\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|, & \text{if } \boldsymbol{A}_{11} \text{ is nonsingular} \\ |\boldsymbol{A}_{22}||\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|, & \text{if } \boldsymbol{A}_{22} \text{ is nonsingular} \end{cases}$$

(h) If $\boldsymbol{A}$ is partitioned as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},$$

where $\boldsymbol{A}_{ij}$ is of order $n_i \times n_j$ $(i,j = 1,2)$, then the inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{B}_{11} &= (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}, \\ \boldsymbol{B}_{12} &= -\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}, \\ \boldsymbol{B}_{21} &= -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11}, \\ \boldsymbol{B}_{22} &= \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}, \end{aligned}$$

provided that the inverses of the matrices that appear in the above four $\boldsymbol{B}_{ij}$ expressions do exist.

The matrix

$$\boldsymbol{S} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \tag{4.25}$$

occurring in Property (g) is called the *Schur complement* of $\boldsymbol{A}_{11}$ in

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$$

when $\boldsymbol{A}_{11}$ is nonsingular. For singular $\boldsymbol{A}_{11}$, the matrix $\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-}\boldsymbol{A}_{12}$ is said to be the *generalized Schur complement* relative to $\boldsymbol{A}_{11}^{-}$ (see Chapter 8). Extensive properties of these complements are to be found in the literature, but this book is not the place for their development. Typical of results, for example, concern inverses of the form

$$(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} = \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}. \tag{4.26}$$

Readers interested in these and other extensions are referred to Marsaglia and Styan(1974a,b), to Henderson and Searle (1981) for results on inverses, and to Ouellette (1981) for general discussion of uses in statistics. All four papers have extensive references.

A further example of the algebra of inverses is a matrix analogue of the scalar result that for scalar $x \neq 1$

$$1 + x + x^2 + \cdots + x^{n-1} = (x^n - 1)/(x - 1).$$

A matrix counterpart is that, provided $(\mathbf{X} - \mathbf{I})^{-1}$ exists

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1} = (\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1}. \tag{4.27}$$

This is established by noting that the product

$$\begin{aligned}
&(\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1})(\mathbf{X} - \mathbf{I}) \\
&\quad = \mathbf{X} + \mathbf{X}^2 + \mathbf{X}^3 + \cdots + \mathbf{X}^{n-1} + \mathbf{X}^n - (\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1}) \\
&\quad = \mathbf{X}^n - \mathbf{I},
\end{aligned} \tag{4.28}$$

and therefore if $(\mathbf{X} - \mathbf{I})^{-1}$ exists, postmultiplication of both sides of (4.28) by $(\mathbf{X} - \mathbf{1})^{-\mathbf{1}}$ yields (4.27). Similarly, it can also be shown that

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \ldots + \mathbf{X}^{n-1} = (\mathbf{X} - \mathbf{I})^{-1}(\mathbf{X}^n - \mathbf{I}).$$

4.13 RANK OF A MATRIX—SOME PRELIMINARY RESULTS

This section discusses an important characteristic of a matrix, namely its rank. Before defining what the rank is, the following theorems are needed:

Theorem 4.1 *A set of linearly independent (LIN) vectors of n elements, each cannot contain more than n such vectors.*

Proof. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be n LIN vectors of order $n \times 1$. Let $\mathbf{u}_{n+1}$ be any other non-null vector of order $n \times 1$. We show that it and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly dependent.

Since $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ has LIN columns, $|\mathbf{U}| \neq \mathbf{0}$ (why?) and $\mathbf{U}^{-1}$ exists. Let $\mathbf{q} = -\mathbf{U}^{-1}\mathbf{u}_{n+1} \neq \mathbf{0}$, because $\mathbf{u}_{n+1} \neq \mathbf{0}$; that is, not all elements of $\mathbf{q}$ are zero. Then $\mathbf{Uq} + \mathbf{u}_{n+1} = \mathbf{0}$, which can be rewritten as

$$q_1\mathbf{u}_1 + q_2\mathbf{u}_2 + \cdots + q_n\mathbf{u}_n + \mathbf{u}_{n+1} = \mathbf{0} \tag{4.29}$$

with not all the q's being zero, This means $\mathbf{u}_1, \mathbf{u}_2, \ldots, , \mathbf{u}_{n+1}$ are linearly dependent; that is, with $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ being LIN it is impossible to put another vector $\mathbf{u}_{n+1}$ with them and have all $n + 1$ vectors being LIN. ■

Corollary 4.1 *When p vectors of order $n \times 1$ are LIN then $p \leq n$.*

It is important to note that this theorem is *not* stating that there is only a single set of n vectors of order $n \times 1$ that are LIN. What it is saying is that if we do have a set of n LIN vectors of order $n \times 1$, then there is no larger set of LIN vectors of order $n \times 1$; that is, there are no sets of $n + 1, n + 2, n + 3, \cdots$ vectors of order $n \times 1$ that are LIN. Although there are many sets of n LIN vectors of order $n \times 1$, an infinite number of them in fact, for each of them it is impossible to put another vector (or vectors) with them and have the set, which then contains more than n vectors, still be LIN.

Example 4.4

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

are a set of two LIN vectors of order 2×1. Put any other vector of order 2×1 with these two vectors and the set will be linearly dependent, that is,

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 7 \\ 8 \end{bmatrix} \quad \textit{are linearly dependent};$$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} a \\ b \end{bmatrix} \quad \textit{are linearly dependent};$$

This latter result is true of course because $\begin{bmatrix} a \\ b \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$. But the theorem is not saying that $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ are the only set of two LIN vectors of order 2×1. For example, $\begin{bmatrix} 5 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 9 \\ 13 \end{bmatrix}$ also form such a set; and any other vector of order 2×1 put with them forms a linearly dependent set of three vectors. This is so because that third vector is a linear combination of these two; for example,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \tfrac{1}{2}(13a - 9b) \begin{bmatrix} 5 \\ 7 \end{bmatrix} + \tfrac{1}{2}(-7a + 5b) \begin{bmatrix} 9 \\ 13 \end{bmatrix}.$$

No matter what the values of a and b are, this expression holds true—that is, every second-order vector can be expressed as a linear combination of a set of two LIN vectors of order 2×1. In general, every nth-*order vector can be expressed as a linear combination of any set of n independent vectors of order $n \times 1$. The maximum number of non-null vectors in a set of independent vectors is therefore n, the order of the vectors. This is simply a restatement of the theorem.*

4.14 THE NUMBER OF LIN ROWS AND COLUMNS IN A MATRIX

Explanation has been given of how a determinant is zero when any of its rows (or columns) are linear combinations of other rows (or columns). In other words, a determinant is zero when its rows (or columns) do not form a set of LIN vectors. Evidently, therefore, a determinant cannot have both its rows forming a dependent set and its columns an independent set, a statement which prompts the more general question of the relationship between the number of LIN rows of a matrix and the number of LIN columns. The relationship is simple.

Theorem 4.2 *The number of LIN rows in a matrix is the same as the number of LIN columns.*

Before proving this, notice that independence of rows (columns) is a property of rows that is unrelated to their sequence within a matrix. For example, because the rows of

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 9 & 14 \\ 3 & 0 & 1 \end{bmatrix} \quad \text{are LIN, so are the rows of} \quad \begin{bmatrix} 6 & 9 & 14 \\ 3 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

Therefore, insofar as general discussion of independence properties is concerned, there is no loss of generality for a matrix that has k LIN rows in assuming that they are the first k rows. This assumption is therefore often made when discussing properties and consequences of independence of rows (columns) of a matrix.

We now prove the theorem.

Proof. Let $\mathbf{A}_{p\times q}$ have k LIN rows and m LIN columns. We show that $k = m$.

Assume that the first k rows $\mathbf{A}$ are LIN and similarly the first m columns. Then partition $\mathbf{A}$ as

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{k\times m} & \mathbf{Y}_{k\times(q-m)} \\ \mathbf{Z}_{(p-k)\times m} & \mathbf{W}_{(p-k)\times(q-m)} \end{bmatrix} \leftarrow k\ \textbf{LIN}\text{ rows.} \tag{4.30}$$

$\uparrow$

m LIN columns

Thus the k rows of $\mathbf{A}$ through $\mathbf{X}$ and $\mathbf{Y}$ are LIN (as are the m columns of $\mathbf{A}$ through $\mathbf{X}$ and $\mathbf{Z}$). Since $\mathbf{A}$ has only k LIN rows the other rows of $\mathbf{A}$ (those through $\mathbf{Z}$ and $\mathbf{W}$) are linear combinations of the first k rows. In particular, the rows of $\mathbf{Z}$ are linear combinations of the rows of $\mathbf{X}$. Hence, these rows can be expressed as $\mathbf{Z} = \mathbf{TX}$ for some matrix $\mathbf{T}$. Now assume that the columns of $\mathbf{X}$ are linearly dependent, that is,

$$\mathbf{Xa} = \mathbf{0} \quad \text{for some vector} \quad \mathbf{a} \neq \mathbf{0}. \tag{4.31}$$

Then $\mathbf{Za} = \mathbf{TXa} = \mathbf{0}$ and so

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \mathbf{a} = \mathbf{0} \quad \text{for that same} \quad \mathbf{a} \neq \mathbf{0}.$$

But this is a statement of the linear dependence of the columns of $\begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix}$, that is, of the first m columns of $\mathbf{A}$. These columns, however, have been taken in (4.30) as being LIN. This is a contradiction, and so assumption (4.31), from which it is derived, is false; that is, the columns of $\mathbf{X}$ are not dependent. Hence they must be LIN.

Having shown the columns of $\mathbf{X}$ to be LIN, observe from (4.30) that there are m of them and they are of order $k \times 1$. Hence by the theorem in the preceding section, $m \leq k$. A similar argument based on the rows of $\mathbf{X}$ and $\mathbf{Y}$, rather than the columns of $\mathbf{X}$ and $\mathbf{Y}$, shows that $k \leq m$. Hence $m = k$. ∎

It is important to notice that this theorem says nothing about which rows (columns) of a matrix are LIN—it is concerned solely with how many of them are LIN. This means, for example, that if there are two LIN rows in a matrix of order 5×4 then once two rows are ascertained as being LIN the other three rows can be expressed as linear combinations of those two. And there may be (there usually is) more than one set of two rows that are LIN. For example, in

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 3 & 2 \\ 3 & 0 & 6 & 5 \\ 2 & 1 & 3 & 3 \\ 6 & 0 & 12 & 10 \end{bmatrix}$$

rows 1 and 2 are LIN, and each of rows 3, 4, and 5 is a linear combination of those first two. Rows 1 and 3 are also LIN, and the remaining rows are linear combinations of them; but not all pairs of rows are LIN. For example, rows 3 and 5 are not linearly independent.

4.15 DETERMINATION OF THE RANK OF A MATRIX

The theorem in Section 4.14 shows that every matrix has the same number of linearly independent rows as it does linearly independent columns.

Definition 4.1 *The* rank *of a matrix is the number of linearly independent rows (and columns) in the matrix.*

Notation: The rank of $\mathbf{A}$ will be denoted equivalently by $r_{\mathbf{A}}$ or $r(\mathbf{A})$. Thus if $r_{\mathbf{A}} \equiv r(\mathbf{A}) = k$, then $\mathbf{A}$ has k LIN rows and k LIN columns. The symbol r is often used for rank, that is, $r_{\mathbf{A}} \equiv r$.

Notice again that no specific set of LIN rows (columns) is identified by knowing the rank of a matrix. Rank indicates only how many are LIN and not where they are located in the matrix. The following properties and consequences of rank are important.

(i) $r_{\mathbf{A}}$ is a positive integer, except that $r_{\mathbf{0}}$ is defined as $r_{\mathbf{0}} = 0$.

(ii) $r(\mathbf{A}_{p\times q}) \leq p$ and $\leq q$: the rank of a matrix equals or is less than the smaller of its number of rows or columns.

(iii) $r(\mathbf{A}_{n\times n}) \leq n$: a square matrix has rank not exceeding its order.

(iv) When $r_{\mathbf{A}} = r \neq 0$ there is at least one square submatrix of $\mathbf{A}$ having order $r \times r$ that is nonsingular. Equation (4.30) with $k = m = r$ is

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{X}_{r\times r} & \mathbf{Y}_{r\times(q-r)} \\ \mathbf{Z}_{(p-r)\times r} & \mathbf{W}_{(p-r)\times(q-r)} \end{bmatrix} \tag{4.32}$$

and $\mathbf{X}_{r\times r}$, the intersection of r linearly independent rows and r linearly columns, is nonsingular. All square submatrices of order greater than $r \times r$ are singular.

(v) When $r(\mathbf{A}_{n\times n}) = n$ then by (iv) $\mathbf{A}$ is nonsingular, that is, $\mathbf{A}^{-1}$ exists. [In (4.32) $\mathbf{A} \equiv \mathbf{X}$.]

(vi) When $r(\mathbf{A}_{n\times n}) < n$ then $\mathbf{A}$ is singular and $\mathbf{A}^{-1}$ does not exist.

(vii) When $r(\mathbf{A}_{p\times q}) = p < q$, $\mathbf{A}$ is said to have *full row rank*, or to be of full row rank. Its rank equals its number of rows.

(viii) When $r(\mathbf{A}_{p\times q}) = q < p$, $\mathbf{A}$ is said to have *full column rank*, or to be of full column rank. Its rank equals its number of columns.

(ix) When $r(\mathbf{A}_{n\times n}) = n$, $\mathbf{A}$ is said to have *full rank*, or to be of full rank. Its rank equals its order, it is nonsingular, its inverse exists, and it is said to be *invertible*.

(x) $r(\boldsymbol{A}) = r(\boldsymbol{A}')$.

(xi) The rank of $\boldsymbol{A}$ does not change if it is pre-multiplied or post-multiplied by a nonsingular matrix. Thus, if $\boldsymbol{A}$ is an $m \times n$ matrix, and $\boldsymbol{B}$ and $\boldsymbol{C}$ are nonsingular matrices of orders $m \times m$ and $n \times n$, respectively, then $r(\boldsymbol{A}) = r(\boldsymbol{BA}) = r(\boldsymbol{AC})$.

(xii) The ranks of $\boldsymbol{AA}'$ and $\boldsymbol{A}'\boldsymbol{A}$ are equal and each is equal to $r(\boldsymbol{A})$.

(xiii) For any matrices, $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k$, having the same number of rows and columns,

$$r\left(\sum_{i=1}^{k} \boldsymbol{A}_i\right) \leq \sum_{i=1}^{k} r(\boldsymbol{A}_i).$$

(xiv) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of orders $m \times n$ and $n \times q$, respectively, then

$$r(\boldsymbol{A}) + r(\boldsymbol{B}) - n \leq r(\boldsymbol{AB}) \leq \min\{r(\boldsymbol{A}), r(\boldsymbol{B})\}$$

This is known as *Sylvester's law*.

(xv) $r(\boldsymbol{A} \otimes \boldsymbol{B}) = r(\boldsymbol{A})r(\boldsymbol{B})$.

(xvi) $r(\boldsymbol{A} \oplus \boldsymbol{B}) = r(\boldsymbol{A}) + r(\boldsymbol{B})$.

Rank is one of the most useful and important characteristics of any matrix. It occurs again and again in this book and plays a vital role throughout all aspects of matrix algebra. For example, from items (vi) and (ix) we see at once that ascertaining whether $|\mathbf{A}|$ is zero or not for determining the existence of $\mathbf{A}^{-1}$ can be replaced by ascertaining whether $r_{\mathbf{A}} < n$ or $r_{\mathbf{A}} = n$. And almost always it is far easier to work with rank than determinants.

TABLE 4.2 Equivalent Statements for the Existence of A^{-1} of Order $n \times n$

Inverse Existing	Inverse Not Existing
$\boldsymbol{A}^{-1}$ exists	$\boldsymbol{A}^{-1}$ does not exist
$\boldsymbol{A}$ is nonsingular	$\boldsymbol{A}$ is singular
$\lvert\boldsymbol{A}\rvert \neq 0$	$\lvert\boldsymbol{A}\rvert = 0$
$\boldsymbol{A}$ has full rank	$\boldsymbol{A}$ has less than full rank
$r_{\boldsymbol{A}} = n$	$r_{\boldsymbol{A}} < n$
$\boldsymbol{A}$ has n linearly independent rows	$\boldsymbol{A}$ has fewer than n linearly independent rows
$\boldsymbol{A}$ has n linearly independent columns	$\boldsymbol{A}$ has fewer than n linearly independent columns
$\boldsymbol{Ax} = \boldsymbol{0}$ has sole solution, $\boldsymbol{x} = \boldsymbol{0}$	$\boldsymbol{Ax} = \boldsymbol{0}$ has many solutions, $\boldsymbol{x} \neq \boldsymbol{0}$

4.16 RANK AND INVERSE MATRICES

A square matrix has an inverse if and only if its rank equals its order. This and other equivalent statements are summarized in Table 4.2. In each half of the table any one of the statements implies all the others: the first and second statements are basically definitional and the last six are equivalences. Hence whenever assurance is needed for the existence of $\mathbf{A}^{-1}$, we need only to establish any one of the last six statements in the first column of the table. The easiest of these is usually that concerning rank: when $r_{\mathbf{A}} = n$ then $\mathbf{A}^{-1}$ exists, and when $r_{\mathbf{A}} < n$ then $\mathbf{A}^{-1}$ does not exist. The problem of ascertaining the existence of an inverse is therefore equivalent to ascertaining if the rank of a square matrix is less than its order.

More generally, there are often occasions when the rank of a matrix is needed exactly. Although locating $r_{\boldsymbol{A}}$ linearly independent rows in $\boldsymbol{A}$ may not always be easy, deriving $r_{\boldsymbol{A}}$ itself is conceptually not difficult.

4.17 PERMUTATION MATRICES

Proof of the theorem in Section 4.14 begins by assuming that all the linearly independent rows come first in a matrix, and the linearly independent columns likewise. But this is not always so. For example, in

$$\boldsymbol{M} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix}$$

there are two linearly independent rows and two linearly independent columns—but not the first two. Now consider $\boldsymbol{E}_{24}\boldsymbol{M}$ for

$$\boldsymbol{E}_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{E}_{24}\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix}. \tag{4.33}$$

Note that $\boldsymbol{E}_{24}$ is an identity matrix with its second and fourth rows interchanged, and $\boldsymbol{E}_{24}\boldsymbol{M}$ is $\boldsymbol{M}$ with those same two rows interchanged.

$\boldsymbol{E}_{24}$ of (4.33) exemplifies what shall be called an *elementary permutation matrix* in general, $\boldsymbol{E}_{rs}$ is an identity matrix with its rth and sth rows interchanged. And $\boldsymbol{E}_{rs}\boldsymbol{M}$ is $\boldsymbol{M}$ with those same two rows interchanged. (The order of $\boldsymbol{E}_{rs}$ is determined by the product it is used in.) By virtue of its definition, $\boldsymbol{E}_{rs}$, is always symmetric. It is also orthogonal (see Section 5.4). Thus $\boldsymbol{E}_{rs}\boldsymbol{E}'_{rs} = \boldsymbol{E}_{rs}\boldsymbol{E}_{rs} = \boldsymbol{I}$ the second equality being true because $\boldsymbol{E}_{rs}\boldsymbol{E}_{rs}$ is $\boldsymbol{E}_{rs}$ with its rth and sth rows interchanged and so is $\boldsymbol{I}$, that is, $\boldsymbol{E}_{rs}\boldsymbol{E}_{rs} = \boldsymbol{I}$.

In the same way that premultiplication of $\boldsymbol{M}$ by $\boldsymbol{E}_{rs}$ interchanges rows r and s of $\boldsymbol{M}$, so does postmultiplication interchange columns. Thus using (4.33),

$$\boldsymbol{E}_{24}\boldsymbol{M}\boldsymbol{E}_{23} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 2 \\ 4 & 12 & 4 \\ 1 & 3 & 1 \end{bmatrix} \tag{4.34}$$

is $\boldsymbol{E}_{24}\boldsymbol{M}$ with its second and third columns interchanged.

Note in (4.34) that $\mathbf{E}_{24}\mathbf{M}\mathbf{E}_{23}$ has its first two rows linearly independent and its first two columns linearly independent also. Hence by premultiplying $\mathbf{M}$ by one elementary permutation matrix and postmultiplying it by another we get a matrix to which the argument of the proof of the theorem applies. And so, because the theorem applies only to the number of linearly independent rows and linearly independent columns, whatever holds for (4.34) in this connection also holds for $\mathbf{M}$.

This use of permutation matrices extends to cases involving more than just a single interchange of rows and/or columns. Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & 3 & 2 \\ 3 & 3 & 9 & 6 \\ 2 & 2 & 5 & 4 \\ 1 & 1 & 7 & 8 \end{bmatrix}. \tag{4.35}$$

In premultiplying $\mathbf{A}$ first by $\mathbf{E}_{34}$ (to interchange rows 3 and 4) and then by $\mathbf{E}_{25}$ (to interchange rows 2 and 5), define $\mathbf{P}$ as $\mathbf{P} = \mathbf{E}_{25}\mathbf{E}_{34}$. Then,

$$\mathbf{PA} = \mathbf{E}_{25}\mathbf{E}_{34}\mathbf{A} = \mathbf{E}_{25}\begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 3 & 2 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 7 & 8 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 3 & 2 \end{bmatrix},$$

where

$$\mathbf{P} = \mathbf{E}_{25}\mathbf{E}_{34} = \mathbf{E}_{25}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then, for $\mathbf{Q} = \mathbf{E}_{24}$,

$$\mathbf{PAQ} = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 8 & 7 & 1 \\ 2 & 4 & 5 & 2 \\ 3 & 8 & 9 & 3 \\ 1 & 2 & 3 & 1 \end{bmatrix}$$

and this is a matrix with its first three rows linearly independent and its first three columns likewise.

In this case $\mathbf{P}$ is a product of two $\mathbf{E}$-matrices. It is an example of a *permutation matrix* which, in general, is an identity matrix with its rows resequenced in some order. Because such a matrix is always a product of elementary permutation matrices (the $\mathbf{E}$-matrices), $\mathbf{P}$ is not necessarily symmetric, but it is always orthogonal—because it is a product of orthogonal $\mathbf{E}$-matrices (see Exercise 9). Furthermore, $\mathbf{P}^{-1} = \mathbf{P}'$ is also a permutation matrix, because although $\mathbf{P}$ is defined as an identity matrix with its rows resequenced, it is also an identity matrix with its columns resequenced. Therefore $\mathbf{P}'$ is an identity matrix with its rows resequenced and so $\mathbf{P}^{-1} = \mathbf{P}'$ is a permutation matrix, too.

One of the great uses of permutation matrices is that for situations like that of the theorem in Section 4.14, permutation matrices provide a mechanism for resequencing rows and columns in a matrix so that a matrix having k linearly independent rows can be resequenced into one having its first k rows and its first k columns linearly independent. Properties of the permutation matrices then allow many properties concerning linear independence of the resequenced matrices to also apply to the original matrix.

4.18 FULL-RANK FACTORIZATION

An immediate consequence of the notion of rank is that a $p \times q$ matrix of rank $r \neq 0$ can be partitioned into a group of r independent rows and a group of $p - r$ rows that are linear combinations of the first group. This leads to a useful factorization.

4.18.1 Basic Development

Equation (4.32) shows the partitioning of $\mathbf{A}_{p\times q}$ of rank r when the first r rows and first r columns are assumed to be linearly independent, and where $\mathbf{X}$ is nonsingular. The first r

rows in (4.32) are those of [**X Y**] and are linearly independent, and so the rows of [**Z W**] are linear combinations of those of [**X Y**]; hence for some matrix **F**,

$$[\mathbf{Z} \quad \mathbf{W}] = \mathbf{F}[\mathbf{X} \quad \mathbf{Y}]. \tag{4.36}$$

Similar reasoning applied to the columns of **A** gives, for some matrix **H**,

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \mathbf{H}. \tag{4.37}$$

But (4.36) has $\mathbf{Z} = \mathbf{FX}$ and $\mathbf{W} = \mathbf{FY}$, and (4.37) has $\mathbf{Y} = \mathbf{XH}$. Hence $\mathbf{W} = \mathbf{FY} = \mathbf{FXH}$, and so therefore

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{XH} \\ \mathbf{FX} & \mathbf{FXH} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \end{bmatrix} \mathbf{X}[\mathbf{I} \quad \mathbf{H}]. \tag{4.38}$$

Furthermore, because X (4.32) is nonsingular, its inverse $\mathbf{X}^{-1}$ exists, so that $\mathbf{Z} = \mathbf{FX}$ and $\mathbf{Y} = \mathbf{XH}$ give $\mathbf{F} = \mathbf{ZX}^{-1}$ and $\mathbf{H} = \mathbf{X}^{-1}\mathbf{Y}$ and then $\mathbf{W} = \mathbf{FY} = \mathbf{ZX}^{-1}\mathbf{Y}$. Hence, (4.38) can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{ZX}^{-1}\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{ZX}^{-1} \end{bmatrix} \mathbf{X}[\mathbf{I} \quad \mathbf{X}^{-1}\mathbf{Y}]. \tag{4.39}$$

Equation (4.38) can be further rewritten in two equivalent ways as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} \\ \mathbf{FX} \end{bmatrix} [\mathbf{I} \quad \mathbf{H}] = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \end{bmatrix} [\mathbf{X} \quad \mathbf{XH}], \tag{4.40}$$

each of which is of the form

$$\mathbf{A}_{p\times q} = \mathbf{K}_{p\times r}\mathbf{L}_{r\times q} \tag{4.41}$$

where **K** has full column rank $r = r_A$, and **L** has full row rank r. We call the matrix product in (4.41) the *full-rank factorization*, after Ben-Israel and Greville(1974, p. 22). It has also been called the *full-rank decomposition* by Marsaglia and Styan (1974a, p. 271). Whatever its name, it can always be done (see Section 4.18.2, which follows) and it has many uses in matrix algebra.

Example 4.5 *Equation (4.39) is illustrated by*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \\ 4 & 9 & 3 & 24 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I}_2 \\ (4 \quad 9)\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 \\ 1 & 14 \end{pmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -3 \\ 0 & 1 & -1 & 4 \end{bmatrix} \tag{4.42}$$

and from (4.40) we have the two equivalent forms of (4.42):

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -3 \\ 0 & 1 & -1 & 4 \end{bmatrix}. \end{aligned} \tag{4.43}$$

A particular case of interest is when $r_{\mathbf{A}} = 1$, whereupon $\mathbf{A} = \mathbf{xy}'$; that is,

$$\begin{bmatrix} 2 & -2 & 4 \\ 3 & -3 & 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} [1 \quad -1 \quad 2].$$

4.18.2 The General Case

Development of (4.36)–(4.40) rests upon the first r rows (and r columns) of $\mathbf{A}$ in (4.32) being linearly independent. But suppose this is not the case, as in (4.35). It is in just such a situation that permutation matrices play their part. Let $\mathbf{M} = \mathbf{PAQ}$ where $\mathbf{P}$ and $\mathbf{Q}$ are permutation matrices, as discussed in Section 4.17, and where $\mathbf{M}$ has the form (4.32). Then $\mathbf{M}$ can be expressed as $\mathbf{M} = \mathbf{KL}$ as in (4.41). Therefore, on using the orthogonality of $\mathbf{P}$ and $\mathbf{Q}$ (that is, $\mathbf{P}^{-1} = \mathbf{P}'$, as at the end of Section 4.17,

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1} = \mathbf{P}'\mathbf{KL}\mathbf{Q}' = (\mathbf{P}'\mathbf{K})(\mathbf{L}\mathbf{Q}') \tag{4.44}$$

where $\mathbf{P}'\mathbf{K}$ and $\mathbf{L}\mathbf{Q}'$ here play the same roles as $\mathbf{K}$ and $\mathbf{L}$ do in (4.41). Hence, (4.41) can be derived quite generally, for any matrix.

4.18.3 Matrices of Full Row (Column) Rank

When $\mathbf{A}_{p\times q}$ has full row rank, there will be no $\mathbf{Z}$ and $\mathbf{W}$ in (4.32); and $\mathbf{K}$ of (4.41) will be an identity matrix. Then (4.32) is $\mathbf{A} = [\mathbf{X}\ \mathbf{Y}]$, and if $\mathbf{X}$ is singular, postmultiplication of $\mathbf{A}$ by a permutation matrix $\mathbf{Q}$ can lead to the partitioning

$$\mathbf{AQ} = [\mathbf{M} \quad \mathbf{L}]$$

where $\mathbf{M}$ is nonsingular. We then have the following lemma.

Lemma 4.1 *A matrix of full row rank can always be written as a product of matrices one of which has the partitioned form $[\mathbf{I}\ \mathbf{H}]$ for some matrix $\mathbf{H}$.*

Proof. From the preceding equations $\mathbf{A} = [\mathbf{M} \quad \mathbf{L}]\mathbf{Q}^{-1} = \mathbf{M}[\mathbf{I} \quad \mathbf{M}^{-1}\mathbf{L}]\mathbf{Q}' = \mathbf{M}[\mathbf{I} \quad \mathbf{H}]\mathbf{Q}'$ for $\mathbf{H} = \mathbf{M}^{-1}\mathbf{L}$, and we have the desired result. ($\mathbf{M}^{-1}$ exists because $\mathbf{M}$ is nonsingular, and $\mathbf{Q}^{-1} = \mathbf{Q}'$ because $\mathbf{Q}$ is a permutation matrix.) ■

REFERENCES

Ben-Israel, A. and Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications*. John Wiley & Sons, New York.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd ed. John Wiley & Sons, New York.

Fisher, R. A. (1949). *The Theory of Inbreeding*. Oliver and Boyd, Edingburgh.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Rev.*, 23, 53–60.

Henderson, H. V., Pukelsheim, F., and Searle, S. R. (1981). On the history of the Kronecker product. Technical Report BU-733-M, Biometrics Unit, Cornell University, Ithaca, NY.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.

Marsaglia, G. and Styan, G. P. H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, 2, 269–292.

Marsaglia, G. and Styan, G. P. H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā*, 36, 437–442.

Ouellette, D. V. (1981). Schur complements and statistics. *Linear Algebra Appl.*, 36, 187–195.

EXERCISES

4.1 For $\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 2 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & -1 & -1 & 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

show that

(a) $\mathbf{AB} = \begin{bmatrix} 3 & -6 & 3 & 12 \\ 2 & -1 & 5 & 5 \end{bmatrix}$ and $\mathbf{A}'\mathbf{B} = \begin{bmatrix} 3 & -2 & 7 & 8 \\ 6 & -1 & 17 & 13 \end{bmatrix}$;

(b) $(\mathbf{A} + \mathbf{A}')\mathbf{B} = \begin{bmatrix} 6 & 8 \\ 8 & 2 \end{bmatrix} \mathbf{B} = \begin{bmatrix} 6 & -8 & 10 & 20 \\ 8 & -2 & 22 & 18 \end{bmatrix} = \mathbf{AB} + \mathbf{A}'\mathbf{B}$;

(c) $\mathbf{BB}' = \begin{bmatrix} 14 & -1 \\ -1 & 3 \end{bmatrix}$ and $\mathbf{B}'\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & 1 & -1 \\ 3 & 1 & 10 & 5 \\ 2 & -1 & 5 & 5 \end{bmatrix}$;

(d) $tr(\boldsymbol{BB}') = tr(\boldsymbol{B}'\boldsymbol{B}) = 17$;

(e) $\mathbf{Bx} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \quad \mathbf{B}'\mathbf{Bx} = \begin{bmatrix} -1 \\ 2 \\ -1 \\ -4 \end{bmatrix}, \quad \mathbf{x}'\mathbf{B}'\mathbf{Bx} = 5 = (\mathbf{Bx})'\mathbf{Bx}$;

(f) $\mathbf{Ay} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \quad \mathbf{A}'\mathbf{Ay} = \begin{bmatrix} -7 \\ -17 \end{bmatrix}, \quad \mathbf{y}'\mathbf{A}'\mathbf{Ay} = 10 = (\mathbf{Ay})'\mathbf{Ay}$;

(g) $\mathbf{A}^2 - 4\mathbf{A} - 9\mathbf{I} = \mathbf{0}$;

(h) $\frac{1}{9}\mathbf{A}\begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix}\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

4.2 Confirm:

(a) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$;

(b) if $\mathbf{A} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then $\mathbf{A}^2 = \mathbf{A}$;

(c) if $\mathbf{B} = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{3} & 1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{3} & 0 \\ 1/\sqrt{6} & 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$, $\mathbf{BB}' = \mathbf{B}'\mathbf{B} = \mathbf{I}_3$;

(d) if $\mathbf{C} = \begin{bmatrix} 6 & -4 \\ 9 & -6 \end{bmatrix}$, then $\mathbf{C}^2$ is null;

(e) $\frac{1}{9}\begin{bmatrix} 4 & -5 & -1 \\ 1 & 1 & 2 \\ 4 & 4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 4 & -1 \end{bmatrix}$ is an identity matrix.

4.3 For $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ -1 & 0 & 7 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 6 & 0 & 0 \\ -3 & 4 & 0 \\ 0 & -5 & 2 \end{bmatrix}$

find $\mathbf{X}^2$, $\mathbf{Y}^2$, $\mathbf{XY}$, and $\mathbf{YX}$, and show that

$$(\mathbf{X} + \mathbf{Y})^2 = \mathbf{X}^2 + \mathbf{XY} + \mathbf{YX} + \mathbf{Y}^2 = \begin{bmatrix} 40 & 5 & 44 \\ -28 & 13 & -33 \\ -1 & -62 & 88 \end{bmatrix}.$$

4.4 Given $\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 0 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & -1 \\ 2 & 3 & 0 \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 6 & 5 & 7 \\ 2 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix}$,

show that $\mathbf{AX} = \mathbf{BX}$ even though $\mathbf{A} \neq \mathbf{B}$.

4.5 **(a)** If, in general $\mathbf{A}$ is $r \times c$, what must be the order of $\mathbf{B}$ so that $\mathbf{A} + \mathbf{B}'$ exists? Why?

(b) Explain why $\mathbf{AVA}' = \mathbf{BVB}'$ implies that $\mathbf{A}$ and $\mathbf{B}$ have the same order and $\mathbf{V}$ is square.

4.6 Under what conditions do both $tr(\mathbf{ABC})$ and $tr(\mathbf{BAC})$ exist? When they exist, do you expect them to be equal? Why? Calculate $tr(\mathbf{ABC})$ and $tr(\mathbf{BAC})$ for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}.$$

4.7 For $\mathbf{A}$ and $\mathbf{B}$ having the same order, explain why

$$\begin{aligned}(\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B})' &= (\mathbf{A}+\mathbf{B})(\mathbf{A}'+\mathbf{B}') \\ &= \mathbf{AA}' + \mathbf{AB}' + \mathbf{BA}' + \mathbf{BB}'.\end{aligned}$$

Will these expressions also generally equal $(\mathbf{A}+\mathbf{B})'(\mathbf{A}+\mathbf{B})$? Why must $\mathbf{A}$ and $\mathbf{B}$ have the same order here? Verify these equalities for

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 1 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 7 \end{bmatrix}.$$

4.8 For a matrix $\mathbf{A}$ from one of the preceding exercises, calculate $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ and verify that the trace of each is the sum of squares of elements of $\mathbf{A}$. Conclude that $tr(A'A) = 0$ if and only if $A = \mathbf{0}$.

4.9 **(a)** When does $(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$?

(b) When $\mathbf{A} = \mathbf{A}'$, prove that $tr(\mathbf{AB}) = tr(\mathbf{AB}')$.

4.10 A generation matrix given by Kempthorne (1957, p. 120) is $\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$. Given that $\mathbf{f}^{(i)} = \mathbf{A}\mathbf{f}^{(i-1)}$, show that

$$\mathbf{f}^{(2)} = \begin{bmatrix} 1 & \frac{3}{4} \\ 0 & \frac{1}{4} \end{bmatrix} \mathbf{f}^{(0)}, \quad \mathbf{f}^{(3)} = \begin{bmatrix} 1 & \frac{7}{8} \\ 0 & \frac{1}{8} \end{bmatrix} \mathbf{f}^{(0)}$$

and $\mathbf{f}^{(n)} = \mathbf{A}^n \mathbf{f}^{(0)} = \begin{bmatrix} 1 & 1 - 1/2^n \\ 0 & 1/2^n \end{bmatrix} \mathbf{f}^{(0)}$.

4.11 With $\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 3 & 1 \\ -4 & 2 & -1 \\ -2 & 0 & 0 \end{bmatrix}$:

(a) Partition $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

where both $\mathbf{A}_{21}$, and $\mathbf{B}_{21}$ have order 1×2.

(b) Calculate $\mathbf{AB}$ both with and without the partitioning, to demonstrate the validity of multiplication of partitioned matrices.

(c) Calculate $\mathbf{AB}'$, showing that

$$\mathbf{B}' = \begin{bmatrix} \mathbf{B}'_{11} & \mathbf{B}'_{21} \\ \mathbf{B}'_{12} & \mathbf{B}'_{22} \end{bmatrix}.$$

4.12 By considering the inner product of $\mathbf{y} - (\mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x})\mathbf{x}$ with itself, prove that

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2.$$

This is the Cauchy–Schwarz inequality. Use it to prove that a product–moment correlation (apart from sign) can never exceed unity.

4.13 Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices of order $n \times n$. Show that the following equalities are equivalent:

(a) $\boldsymbol{AA}'\boldsymbol{BB}' = \boldsymbol{AB}'\boldsymbol{AB}'$.

(b) $tr(\boldsymbol{AA}'\boldsymbol{BB}') = tr(\boldsymbol{AB}'\boldsymbol{AB}')$

(c) $\boldsymbol{A}'\boldsymbol{B} = \boldsymbol{B}'\boldsymbol{A}$.

4.14 **(a)** Let $\boldsymbol{A}$ be a matrix of order $n \times n$. Show that the sum of all elements of $\boldsymbol{A}$ is equal to $\mathbf{1}_n'\boldsymbol{A}\mathbf{1}_n$, where $\mathbf{1}_n$ is a vector whose n elements are all equal to one.

(b) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of order $n \times n$. Show that $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same sum of diagonal elements.

(Hint: Use the vector $\boldsymbol{u}_i$ whose elements are all equal to zero except for the ith element which is equal to one, $i = 1, 2, \ldots, n$.)

4.15 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of order $n \times n$ such that $\boldsymbol{A} = \boldsymbol{A}^2$ and $\boldsymbol{B} = \boldsymbol{B}^2$. Show that if $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{B}$, then $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{B}$.

4.16 Let $\boldsymbol{A}$ be a square matrix of order $n \times n$ and $\mathbf{1}$ be a vector of ones of order $n \times 1$. Suppose that $\boldsymbol{x} \neq \mathbf{0}$ exists such that $\boldsymbol{Ax} = \mathbf{0}$ and $\boldsymbol{x}'\mathbf{1} = 0$. Prove that

(a) $\boldsymbol{A} + \lambda\mathbf{1}\mathbf{1}'$ is singular for any scalar λ,

(b) $\boldsymbol{A} + \boldsymbol{f}\mathbf{1}'$ is also singular for any vector $\boldsymbol{f}$.

4.17 Let $\boldsymbol{X}$ be a matrix of order $n \times p$ and rank p $(n \geq p)$. What is the rank of $\boldsymbol{XX}'\boldsymbol{X}$?

4.18 Let $\boldsymbol{X}$ be a matrix of order $n \times n$ that satisfies the equation

$$\boldsymbol{X}^2 + 2\boldsymbol{X} + \boldsymbol{I}_n = \mathbf{0}.$$

(a) Show that $\boldsymbol{X} + \boldsymbol{I}_n$ is singular.

(b) Show that $\boldsymbol{X}$ is nonsingular.

(c) Show that $\boldsymbol{X} + 2\boldsymbol{I}_n$ is nonsingular.

(d) Derive an expression for $\boldsymbol{X}^{-1}$.

4.19 Prove that $\boldsymbol{vv}' - \boldsymbol{v}'\boldsymbol{vI}$ is singular.

4.20 Show that if

(a) $\mathbf{A} = \begin{bmatrix} 6 & 13 \\ 5 & 12 \end{bmatrix}$, then $|\mathbf{A}| = 7$, $\mathbf{A}^{-1} = \frac{1}{7}\begin{bmatrix} 12 & -13 \\ -5 & 6 \end{bmatrix}$

and $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$;

(b) $\mathbf{B} = \begin{bmatrix} 3 & -4 \\ 7 & 14 \\ & \end{bmatrix}$, then $|\mathbf{B}| = 70$, $\mathbf{B}^{-1} = \frac{1}{70}\begin{bmatrix} 14 & 4 \\ -7 & 3 \\ & \end{bmatrix}$

and $\mathbf{BB}^{-1} = \mathbf{B}^{-1} = \mathbf{B} = \mathbf{I}$

4.21 Demonstrate the reversal rule for the inverse of a product of two matrices, using **A** and **B** given in Exercise 20.

4.22 Show that if

(a) $\mathbf{A} = \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 4 \end{bmatrix}$, $|\mathbf{A}| = -5$, $\mathbf{A}^{-1} = \frac{1}{5}\begin{bmatrix} -8 & 12 & -3 \\ -1 & 4 & -1 \\ 2 & -3 & 2 \end{bmatrix}$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}$;

(b) $\mathbf{B} = \begin{bmatrix} 10 & 6 & -1 \\ 6 & 5 & 4 \\ -1 & 4 & 17 \end{bmatrix}$, $|\mathbf{B}| = 25$, $\mathbf{B}^{-1} = \begin{bmatrix} 2.76 & -4.24 & 1.16 \\ -4.24 & 6.76 & -1.84 \\ 1.16 & -1.84 & 0.56 \end{bmatrix}$

and $\mathbf{B}^{-1}\mathbf{B} = \mathbf{BB}^{-1} = \mathbf{I}$;

(c) $\mathbf{C} = \frac{1}{10}\begin{bmatrix} 0 & -6 & 8 \\ -10 & 0 & 0 \\ 0 & -8 & -6 \\ & & \end{bmatrix}$, then $|\mathbf{C}| = 1$, $\mathbf{C}^{-1} = \mathbf{C}'$ and $\mathbf{CC}' = \mathbf{C}'\mathbf{C} = \mathbf{I}$.

4.23 For $\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{ww}'$ with $\boldsymbol{w}'\boldsymbol{w} = 1$, prove that $\boldsymbol{H} = \boldsymbol{H}^{-1}$.

5

Special Matrices

Numerous matrices with particular properties have attracted special names. Although their historical origins are in specific mathematical problems or applications, they were later found to have properties of broader interest. Of the vast number of such matrices, this chapter presents but a small selection of special matrices that are commonly used in statistics. Other matrices appear in subsequent chapters.

5.1 SYMMETRIC MATRICES

Definition 5.1 *A square matrix is defined as symmetric when it equals its transpose; that is,*

$$\mathbf{A} \textit{ is symmetric when } \mathbf{A} = \mathbf{A}', \textit{ with } a_{ij} = a_{ji} \tag{5.1}$$

for $i, j = \mathrm{i}, \ldots, r$ *for* $\mathbf{A}_{r \times r}$.

Symmetric matrices have many useful special properties, a few of which (relating to matrix products) are noted here. Others are noted in later chapters.

5.1.1 Products of Symmetric Matrices

Products of symmetric matrices are not generally symmetric. If $\mathbf{A} = \mathbf{A}'$ and $\mathbf{B} = \mathbf{B}'$, then $\mathbf{A}$ and $\mathbf{B}$ are symmetric but the transpose of their product (when it exists) is

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' = \mathbf{BA}.$$

Since $\mathbf{BA}$ is generally not the same as $\mathbf{AB}$, this means $\mathbf{AB}$ is generally not symmetric.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

Example 5.1 *With*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \textit{and} \quad \mathbf{B} = \begin{bmatrix} 3 & 7 \\ 7 & 6 \end{bmatrix},$$

$$(\mathbf{AB})' = \begin{bmatrix} 17 & 19 \\ 27 & 32 \end{bmatrix}' = \begin{bmatrix} 17 & 27 \\ 19 & 32 \end{bmatrix} = \mathbf{BA} \neq \mathbf{AB}.$$

5.1.2 Properties of AA′ and A′A

Products of a matrix and its transpose always exist and are symmetric:

$$(\mathbf{AA}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{AA}' \quad \text{and} \quad (\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}. \tag{5.2}$$

Observe the method used for showing that a matrix is symmetric: transpose the matrix and show that the result equals the matrix itself.

Although both products $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are symmetric, they are not necessarily equal. In fact, only when $\mathbf{A}$ is square might they be equal, because it is only then that both products have the same order. Thus for $\mathbf{A}_{r\times c}$, the product $\mathbf{AA}' = \mathbf{A}_{r\times c}(\mathbf{A}')_{c\times r}$ has order $r \times r$, and $\mathbf{A}'\mathbf{A} = (\mathbf{A}')_{c\times r}\mathbf{A}_{r\times c}$ has order $c \times c$.

Example 5.2 *For*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{bmatrix},$$

$$\mathbf{AA}' = \begin{bmatrix} 6 & 2 \\ 2 & 10 \end{bmatrix}, \quad \textit{and} \quad \mathbf{A}'\mathbf{A} = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 4 & -2 \\ 2 & -2 & 2 \end{bmatrix}.$$

Note that matrix multiplication ensures that elements of $\mathbf{AA}'$ are inner products of rows of $\mathbf{A}$ with themselves and with each other:

$$\mathbf{AA}' = \{\text{inner product of } i\text{th and } k\text{th rows of } \mathbf{A}\} \text{ for } i, k = 1, \ldots, r. \tag{5.3}$$

In particular, the ith diagonal element of $\mathbf{AA}'$ is the sum of squares of the elements of the ith row of $\mathbf{A}$, namely $\sum_{j=1}^{c} a_{ij}^2$ for $\mathbf{A}$ of order $r \times c$. Confining attention to real matrices, we can use the property of real numbers that a sum of squares of them is positive (unless they are all zero), and so observe that for real matrices $\mathbf{A}$, the matrix $\mathbf{AA}'$ has diagonal elements that are positive (or zero). Similar results hold for $\mathbf{A}'\mathbf{A}$ in terms of columns of $\mathbf{A}$:

$$\mathbf{A}'\mathbf{A} = \{\text{inner product of } j\text{th } \textit{and } m\text{th } \textit{columns of } \mathbf{A}\} \text{ for } j, m = 1, \ldots, c. \tag{5.4}$$

The reader can verify these results for the example.

Recall that if a sum of squares of real numbers is zero, then each of the numbers is zero; that is, for real numbers $x_1, x_2, \ldots, x_n$, $\Sigma_{i=1}^n x_i^2 = 0$ implies $x_1 = 0 = x_2 = \cdots = x_n$. This is the basis for proving, for any real matrix $\mathbf{A}$, that

$$\mathbf{A}'\mathbf{A} = \mathbf{0} \quad \text{implies} \quad \mathbf{A} = \mathbf{0} \tag{5.5}$$

and

$$tr(\mathbf{A}'\mathbf{A}) = \mathbf{0} \quad \text{implies} \quad \mathbf{A} = \mathbf{0}. \tag{5.6}$$

Result (5.5) is true because $\mathbf{A}'\mathbf{A} = \mathbf{0}$ implies that every diagonal element of $\mathbf{A}'\mathbf{A}$ is zero; and for the jth such element this means, by (5.3), that the inner product of the jth column of $\mathbf{A}$ with itself is zero; that is, for $\mathbf{A}$ of order $r \times c$, that $\sum_{k=1}^r a_{kj}^2 = 0$ for $j = 1, \ldots, c$. Hence $a_{kj} = 0$ for $k = 1, \ldots, r$; and this is true for $j = 1, \ldots, c$. Therefore, $\mathbf{A} = \mathbf{0}$. Proof of (5.6) is similar:

$$tr(\mathbf{A}'\mathbf{A}) = \sum_{j=1}^{c} (j\text{th diagonal element of } \mathbf{A}'\mathbf{A}) = \sum_{j=1}^{c} \sum_{k=1}^{r} a_{kj}^2$$

and so $tr(\mathbf{A}'\mathbf{A}) = \mathbf{0}$ implies that every a_{kj} is zero; that is, every element of $\mathbf{A}$ is zero. Hence $\mathbf{A} = \mathbf{0}$.

Results (5.5) and (5.6) are seldom useful for the sake of some particular matrix $\mathbf{A}$, but they are often helpful in developing other results in matrix algebra when $\mathbf{A}$ is a function of other matrices. For example, by means of (5.5) we can prove, for real matrices $\mathbf{P}, \mathbf{Q}$, and $\mathbf{X}$, that

$$\mathbf{PXX}' = \mathbf{QXX}' \quad \text{implies} \quad \mathbf{PX} = \mathbf{QX}. \tag{5.7}$$

The proof consists of observing that

$$(\mathbf{PXX}' - \mathbf{QXX}')(\mathbf{P}' - \mathbf{Q}') \equiv (\mathbf{PX} - \mathbf{QX})(\mathbf{PX} - \mathbf{QX})'. \tag{5.8}$$

Hence if $\mathbf{PXX}' = \mathbf{QXX}'$, the left-hand side of (5.8) is null and so, therefore, is the right-hand side; hence by (5.5), $\mathbf{PX} - \mathbf{QX} = \mathbf{0}$; that is, $\mathbf{PX} = \mathbf{QX}$, and (5.7) is established.

5.1.3 Products of Vectors

The inner product of two vectors is a scalar and is therefore symmetric: $\mathbf{x}'\mathbf{y} = (\mathbf{x}'\mathbf{y})' = \mathbf{y}'\mathbf{x}$. In contrast, the outer product of two vectors (see Section 4.7.5) is not necessarily symmetric: $\mathbf{xy}' = (\mathbf{yx}')' \neq (\mathbf{xy}')'$. Indeed, such a product is not necessarily even square.

Example 5.3 *For* $\mathbf{x}' = [1 \quad 0 \quad 2]$, $\mathbf{y}' = [2 \quad 2 \quad 1]$, *and* $\mathbf{z}' = [3 \quad 2]$,

$$\mathbf{xy}' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 0 \\ 4 & 4 & 2 \end{bmatrix} \tag{5.9}$$

is not symmetric. Neither is

$$\mathbf{yx}' = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 0 & 4 \\ 1 & 0 & 2 \end{bmatrix}, \tag{5.10}$$

but its transpose does, of course, equal $\mathbf{xy}'$*; that is, (5.10) equals the transpose of (5.9); namely* $\mathbf{yx}' = (\mathbf{xy}')'$. *And* $\mathbf{xz}'$ *and* $\mathbf{zx}'$ *are rectangular:*

$$\mathbf{xz}' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 0 \\ 6 & 4 \end{bmatrix}$$

and

$$\mathbf{zx}' = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 \\ 2 & 0 & 4 \end{bmatrix}.$$

5.1.4 Sums of Outer Products

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_c$ be the columns of $\boldsymbol{A}$, and $\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \ldots, \boldsymbol{\beta}'_c$ be the rows of $\boldsymbol{B}$, then the product $\boldsymbol{AB}$ expressed as,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \cdots \mathbf{a}_c \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}'_1 \\ \boldsymbol{\beta}'_2 \\ \vdots \\ \boldsymbol{\beta}'_c \end{bmatrix} = \sum_{J=1}^{c} \mathbf{a}_j \boldsymbol{\beta}'_j, \tag{5.11}$$

is the sum of outer products of the columns of $\mathbf{A}$ with the corresponding rows in $\mathbf{B}$.

Example 5.4

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} = \begin{bmatrix} 43 & 48 \\ 59 & 66 \\ 75 & 84 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 7 & 8 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \begin{bmatrix} 9 & 10 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ 14 & 16 \\ 21 & 24 \end{bmatrix} + \begin{bmatrix} 36 & 40 \\ 45 & 50 \\ 54 & 60 \end{bmatrix}. \end{aligned}$$

A special case of (5.11) is when $\mathbf{B} = \mathbf{A}'$*:*

$$\mathbf{AA}' = \sum_{j=1}^{c} \mathbf{a}_j \mathbf{a}'_j. \tag{5.12}$$

$\mathbf{AA}'$ *is thus the sum of outer products of each column of* $\mathbf{A}$ *with itself; that is,*

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}\begin{bmatrix} 1 & 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix}\begin{bmatrix} 2 & 5 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \end{bmatrix}\begin{bmatrix} 3 & 6 \end{bmatrix}.$$

5.1.5 Elementary Vectors

A special case of (5.12) is

$$\mathbf{I}_n = \sum_{i=1}^{n} \mathbf{e}_i \mathbf{e}_i'$$

for $\mathbf{e}_i$ being the ith column of $\mathbf{I}_n$, namely a vector with unity for its ith element and zeros elsewhere. $\mathbf{e}_i$ is called an *elementary vector*. When necessary, its order n can be identified by denoting $\mathbf{e}_i$ as $\mathbf{e}_i^{(n)}$.

The outer product of one elementary vector with another is a null matrix except for one element of unity: for example, with

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{E}_{12} = \mathbf{e}_1 \mathbf{e}_2' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In general, $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j'$ is null except for element (i, j) being unity; and, of course, $\mathbf{I} = \Sigma_i \mathbf{E}_{ii}$. These $\mathbf{E}_{ij}$-matrices are particularly useful in applications of calculus to matrix algebra (see Section 9.4.3)

The $\mathbf{e}$-vectors are also useful for delineating individual rows and columns of a matrix. Thus $\mathbf{e}_i'\mathbf{A} = \alpha_i'$, the ith row of $\mathbf{A}$ and $\mathbf{A}\mathbf{e}_j = \mathbf{a}_j$, the jth column of $\mathbf{A}$; that is, for

$$\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}, \quad \mathbf{e}_2'\mathbf{A} = \begin{bmatrix} 5 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{e}_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

5.1.6 Skew-Symmetric Matrices

A symmetric matrix $\mathbf{A}$ has the property $\mathbf{A} = \mathbf{A}'$; in contrast there are also matrices $\mathbf{B}$ having the property $\mathbf{B}' = -\mathbf{B}$. Their diagonal elements are zero and each off-diagonal element is minus its symmetric partner; that is, $b_{ii} = 0$ and $b_{ij} = -b_{ji}$. An example is

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 3 & -2 & 0 \end{bmatrix}.$$

Such matrices, having $\mathbf{B}' = -\mathbf{B}$, are called *skew-symmetric*.

5.2 MATRICES HAVING ALL ELEMENTS EQUAL

Vectors whose every element is unity are called *summing vectors* because they can be used to express a sum of numbers in matrix notation as an inner product.

Example 5.5 *The row vector* $\mathbf{1}' = [1 \quad 1 \quad 1 \quad 1]$ *is the summing vector of order* 1×4 *and for* $\mathbf{x}' = [3 \quad 6 \quad 8 \quad -2]$

$$\mathbf{1}'\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \\ 8 \\ -2 \end{bmatrix} = 3 + 6 + 8 - 2$$

$$= 15 = \begin{bmatrix} 3 & 6 & 8 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{x}'\mathbf{1}.$$

When necessary to avoid confusion, the order of a summing vector can be denoted in the usual way: $\mathbf{1}'_4 = [1 \quad 1 \quad 1 \quad 1]$. *For example,*

$$\mathbf{1}'_3\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -5 & -3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} = \mathbf{1}'_2.$$

Were this equation in $\mathbf{X}$ *to be written as* $\mathbf{1}'\mathbf{X} = \mathbf{1}'$, *one might be tempted to think that both* $\mathbf{1}'$ *vectors had the same order. That they do not is made clear by denoting* $\mathbf{X}$ *of order* $r \times c$ *as* $\mathbf{X}_{r\times c}$ *so that* $\mathbf{1}'\mathbf{X}_{r\times c} = \mathbf{1}'$ *is* $\mathbf{1}'_r\mathbf{X}_{r\times c} = \mathbf{1}'_c$.

The inner product of a summing vector with itself is a scalar, the vector's order:

$$\mathbf{1}'_n\mathbf{1}_n = n. \tag{5.13}$$

And outer products are matrices with all elements unity:

$$\mathbf{1}_3\mathbf{1}'_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{J}_{3\times 2}.$$

In general, $\mathbf{1}_r\mathbf{1}'_s$ has order $r \times s$ and is often denoted by the symbol $\mathbf{J}$ or $\mathbf{J}_{r\times s}$:

$$\mathbf{1}_r\mathbf{1}'_s = \mathbf{J}_{r\times s}, \ \text{having all elements unity.} \tag{5.14}$$

Clearly, $\lambda\mathbf{J}_{r\times s}$ has all elements λ.

Products of $\mathbf{J}$'s with each other and with $\mathbf{1}$'s are, respectively, $\mathbf{J}$'s and $\mathbf{1}$'s (multiplied by scalars):

$$\mathbf{J}_{r\times s}\mathbf{J}_{s\times t} = s\mathbf{J}_{r\times t}, \qquad \mathbf{1}'_r\mathbf{J}_{r\times s} = r\mathbf{1}'_s \quad \text{and } \mathbf{J}_{r\times s}\mathbf{1}_s = s\mathbf{1}_r. \tag{5.15}$$

Particularly useful are square **J**'s and a variant thereof:

$$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}'_n \quad \text{with} \quad \mathbf{J}_n^2 = n\mathbf{J}_n;$$

and (5.16)

$$\bar{\mathbf{J}}_n = \frac{1}{n}\mathbf{J}_n \quad \text{with} \quad \bar{\mathbf{J}}_n^2 = \bar{\mathbf{J}}_n.$$

And for statistics

$$\mathbf{C}_n = \mathbf{I} - \bar{\mathbf{J}}_n = \mathbf{I} - \frac{1}{n}\mathbf{J}_n, \tag{5.17}$$

known as a *centering matrix*, is especially useful, as is now illustrated. First observe that

$$\mathbf{C} = \mathbf{C}' = \mathbf{C}^2, \qquad \mathbf{C1} = \mathbf{0} \quad \text{and} \quad \mathbf{CJ} = \mathbf{JC} = \mathbf{0}, \tag{5.18}$$

which the reader can easily verify.

Example 5.6 *The mean and sum of squares about the mean for data $x_1, x_2, \dots, x_n$ can be expressed in terms of **1**-vectors and **J**-matrices. Define*

$$\mathbf{x}' = [x_1 \quad x_2 \quad \cdots \quad x_n].$$

Then the mean of the x's is

$$\bar{x} = (x_1 + x_2 + \cdots + x_n)/n = \sum_{i=1}^{n} x_i/n = \frac{1}{n}\mathbf{x}'\mathbf{1} = \frac{1}{n}\mathbf{1}'\mathbf{x},$$

*the last equality arising from $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$. Using **C** of (5.17) and (5.18), we get*

$$\mathbf{x}'\mathbf{C} = \mathbf{x}' - \mathbf{x}'\bar{\mathbf{J}} = \mathbf{x}' - \frac{1}{n}\mathbf{x}'\mathbf{11}' = \mathbf{x}' - \bar{\mathbf{x}}\mathbf{1}' = \{\mathbf{x}_i - \bar{\mathbf{x}}\}$$

*is the data vector with each observation expressed as a deviation from $\bar{x}$. (This is the origin of the name centering matrix for **C**.) Postmultiplying $\mathbf{x}'\mathbf{C}$ by $\mathbf{x}$ then gives*

$$\mathbf{x}'\mathbf{Cx} = (\mathbf{x}' - \bar{\mathbf{x}}\mathbf{1}')\mathbf{x} = \mathbf{x}'\mathbf{x} - \bar{\mathbf{x}}(\mathbf{1}'\mathbf{x}) = \mathbf{x}'\mathbf{x} - n\bar{x}^2.$$

Hence, using a standard result in statistics, we get

$$\sum_{\imath-1}^{n}(x_\imath - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \mathbf{x}'\mathbf{x} - n\bar{x}^2 = \mathbf{x}'\mathbf{Cx}. \tag{5.19}$$

Thus for $\mathbf{x}'$ being a data vector, $\mathbf{x}'\mathbf{1}/n$ is the mean, $\mathbf{x}'\mathbf{C}$ is the vector of deviations from the mean, and $\mathbf{x}'\mathbf{Cx}$ is the sum of squares about the mean.

Expression (5.19) is a special case of the form $\mathbf{x}'\mathbf{Ax}$, known as a quadratic form (see Section 5.6), which can be used for sums of squares generally. Expressed in this manner, and with the aid of other matrix concepts (notably idempotency; see Section 5.3), sums of squares of normally distributed x's are known to be distributed as χ^2 under very simply stated (in matrix notation) conditions (see Section 10.5).

Example 5.7 *It was noted in Example 2.1 that the row sums of a transition probability matrix are always unity [see formula (2.6)]. For example, with*

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix}, \qquad \begin{bmatrix} 0.2 + 0.8 \\ 0.4 + 0.6 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \text{ i.e.,} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This last result is

$$\mathbf{P1} = \mathbf{1},$$

which is true generally for any transition probability matrix. Furthermore, because $\mathbf{P}^2\mathbf{1} = \mathbf{P}(\mathbf{P1}) = \mathbf{P1} = \mathbf{1}$ *this in turn extends to*

$$\mathbf{P}^n\mathbf{1} = \mathbf{1},$$

showing that row sums of any power of a transition probability matrix are also unity.

5.3 IDEMPOTENT MATRICES

The matrix $\bar{\mathbf{J}}_n$ in (5.17) has the characteristic that its square equals itself. Many different matrices are of this nature; they are called idempotent matrices. Thus when $\mathbf{K}$ is such that $\mathbf{K}^2 = \mathbf{K}$, we say $\mathbf{K}$ is *idempotent* (from Latin, idem meaning "same," and potent "power"). All idempotent matrices are square (otherwise $\mathbf{K}^2$ does not exist); identity matrices and square null matrices are idempotent. When $\mathbf{K}$ is idempotent, all powers of $\mathbf{K}$ equal $\mathbf{K}$; that is, $\mathbf{K}^r = \mathbf{K}$ for r being a positive integer, and $(\mathbf{I} - \mathbf{K})$ is idempotent. Thus

$$\mathbf{K}^2 = \mathbf{K} \text{ implies } (\mathbf{I} - \mathbf{K})^2 = \mathbf{I} - \mathbf{K},$$

but $\mathbf{K} - \mathbf{I}$ is not idempotent. A product of two idempotent matrices is idempotent if the matrices commute in multiplication.

Idempotent matrices occur in many applications of matrix algebra and they play an especially important role in statistics.

Example 5.8 *The matrix* $\mathbf{I} - \bar{\mathbf{J}}$ *in (5.17) is idempotent; so is* $\mathbf{GA}$ *whenever* $\mathbf{G}$ *is such that* $\mathbf{AGA} = \mathbf{A}$ [*because then* $(\mathbf{GA})^2 = \mathbf{GAGA} = \mathbf{GA}$]. *A matrix* $\mathbf{G}$ *of this nature is called a generalized inverse of* $\mathbf{A}$*; its properties are considered in Chapter 8.*

A matrix $\mathbf{A}$ satisfying $\mathbf{A}^2 = \mathbf{0}$ is called *nilpotent*, and that for which $\mathbf{A}^2 = \mathbf{I}$ could be called *unipotent*.

Example 5.9

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \textit{ is nilpotent}; \quad \mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{X} \\ 0 & -\mathbf{I} \end{bmatrix} \textit{ is unipotent}.$$

Variations on these definitions are $\mathbf{A}^k = \mathbf{A}$, $\mathbf{A}^k = \mathbf{0}$, and $\mathbf{A}^k = \mathbf{I}$ for some positive integer $k > 2$. An example is the matrix

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix},$$

for which $\mathbf{B}^3 = \mathbf{I}$, but $\mathbf{B}^2 \neq \mathbf{I}$.

The following theorem is very useful in determining the rank of an idempotent matrix:

Theorem 5.1 *If $\boldsymbol{K}$ is an $n \times n$ idempotent matrix, then its rank r is equal to its trace, that is, $r = r(\boldsymbol{K}) = tr(\boldsymbol{K})$.*

Proof. Let $\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_r$ be linearly independent vectors that span (form a basis for) the column space of $\boldsymbol{K}$. Let $\boldsymbol{L} = [l_1 : l_2 : \ldots : \boldsymbol{l}_r]$, then $\boldsymbol{L}$ is of order $n \times r$ and rank r. The ith column, k_i of $\boldsymbol{K}$ can then be expressed as a linear combination of the columns of $\boldsymbol{L}$ $(i = 1, 2, \ldots, n)$. We can therefore write $\boldsymbol{k}_i = \boldsymbol{L}\boldsymbol{m}_i$ where $\boldsymbol{m}_i$ is a vector of coefficients consisting of r elements $(i = 1, 2, \ldots, n)$. Let $\boldsymbol{M}$ be a matrix of order $r \times n$ whose columns are $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_n$. Thus $\boldsymbol{K}$ can be written as

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{M}. \tag{5.20}$$

It follows that $r(\boldsymbol{K}) \leq r(\boldsymbol{M}) \leq r$, since the rank of $\boldsymbol{M}$ cannot exceed the number of its rows. But, $r(\boldsymbol{K}) = r$. We conclude that $r(\boldsymbol{M}) = r$.

Since $\boldsymbol{K}$ is idempotent, $\boldsymbol{K}^2 = \boldsymbol{K}$. We then have from (5.20),

$$\boldsymbol{L}\boldsymbol{M}\boldsymbol{L}\boldsymbol{M} = \boldsymbol{L}\boldsymbol{M}. \tag{5.21}$$

Furthermore, because $\boldsymbol{L}$ is of full-column rank, multiplying both sides of (5.21) on the left by $\boldsymbol{L}'$ and noting that $\boldsymbol{L}'\boldsymbol{L}$ is nonsingular by the fact that it is of order $r \times r$ of rank r, we get, after multiplying both of the resulting sides of (5.21) on the left by the inverse of $\boldsymbol{L}'\boldsymbol{L}$,

$$\boldsymbol{M}\boldsymbol{L}\boldsymbol{M} = \boldsymbol{M}. \tag{5.22}$$

Similarly, $\boldsymbol{M}$ being of full-row rank, the matrix $\boldsymbol{M}\boldsymbol{M}'$ is nonsingular. Multiplying the two sides of (5.22) on the right by $\boldsymbol{M}'$ and then multiplying the resulting equation on the right by the inverse of $\boldsymbol{M}\boldsymbol{M}'$, we get

$$\boldsymbol{M}\boldsymbol{L} = \boldsymbol{I}_r. \tag{5.23}$$

It follows from (5.20) and (5.23) that $r(\boldsymbol{K}) = r = tr(\boldsymbol{ML}) = tr(\boldsymbol{LM}) = tr(\boldsymbol{K})$. Thus, $r(\boldsymbol{K}) = tr(\boldsymbol{K})$. ■

Since the trace of $\boldsymbol{K}$, being the sum of its diagonal elements, is easy to compute, this theorem facilitates the finding of the rank of $\boldsymbol{K}$, which, in general, is more difficult to determine, when $\boldsymbol{K}$ is idempotent.

5.4 ORTHOGONAL MATRICES

Definition 5.2 *Another useful class of matrices is that for which* $\mathbf{A}$ *has the property* $\mathbf{AA}' = \mathbf{I} = \mathbf{A}'\mathbf{A}$. *Such matrices are called* orthogonal. *We lead up to them with the following definitions*:

The norm of a real vector $\mathbf{x}' = [x_1 \quad x_2 \quad \cdots \quad x_n]$ *is defined as*

$$\text{norm of } \mathbf{x} = \sqrt{\mathbf{x}'\mathbf{x}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}} \tag{5.24}$$

For example, the norm of $\mathbf{x}' = [1 \quad 2 \quad 2 \quad 4]$ is $(1 + 4 + 4 + 16)^{\frac{1}{2}} = 5$. (The square root is taken as positive.) A vector is said to be either *normal* or a *unit vector* when its norm is unity; that is, when $\mathbf{x}'\mathbf{x} = 1$. An example is $\mathbf{x}' = [.2 \quad .4 \quad .4 \quad .8]$. Any non-null vector $\mathbf{x}$ can be changed into a unit vector by multiplying it by the scalar $1/\sqrt{\mathbf{x}'\mathbf{x}}$; that is,

$$\mathbf{u} = \left(\frac{1}{\sqrt{\mathbf{x}'\mathbf{x}}} \right) \mathbf{x}$$

is the *normalized* form of $\mathbf{x}$ (because $\mathbf{u}'\mathbf{u} = 1$).

Non-null vectors $\mathbf{x}$ and $\mathbf{y}$ are described as being *orthogonal* when $\mathbf{x}'\mathbf{y} = 0$ (equivalent, of course, to $\mathbf{y}'\mathbf{x} = 0$); for example, $\mathbf{x}' = [1 \quad 2 \quad 2 \quad 4]$ and $\mathbf{y}' = [6 \quad 3 \quad -2 \quad -2]$ are orthogonal vectors because

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} 1 & 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \\ -2 \\ -2 \end{bmatrix} = 6 + 6 - 4 - 8 = 0.$$

Two vectors are defined as *orthonormal vectors* when they are orthogonal and normal. Thus $\mathbf{u}$ and $\mathbf{v}$ are orthonormal when $\mathbf{u}'\mathbf{u} = 1 = \mathbf{v}'\mathbf{v}$ and $\mathbf{u}'\mathbf{v} = 0$; for example, $\mathbf{u}' = \frac{1}{6}[1 \quad 1 \quad 3 \quad 3 \quad 4]$ and $\mathbf{v}' = [-0.1 \quad -0.9 \quad -0.1 \quad -0.1 \quad 0.4]$ are orthonormal vectors.

A group, or collection, of vectors all of the same order is called a *set* of vectors. A set of vectors $\mathbf{x}_i$ for $i = 1, 2, \ldots, n$ is said to be an orthonormal set of vectors when every vector in the set is normal, $\mathbf{x}_i'\mathbf{x}_i = 1$ for all i, and when every pair of different vectors in the set is orthogonal, $\mathbf{x}_i'\mathbf{x}_j = 0$ for $i \neq j = 1, 2, \ldots, n$. We can say that the vectors of an orthonormal set are all normal, and pairwise orthogonal.

A matrix $\mathbf{P}_{r\times c}$ whose rows constitute an orthonormal set of vectors is said to have orthonormal rows, whereupon $\mathbf{PP}' = \mathbf{I}_r$. But then $\mathbf{P}'\mathbf{P}$ is not necessarily an identity matrix $\mathbf{I}_c$, as the following example shows:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{PP}' = \mathbf{I}_2 \quad \text{but} \quad \mathbf{P}'\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \mathbf{I}_2.$$

Conversely, when $\mathbf{P}_{r\times c}$ has orthonormal columns $\mathbf{P}'\mathbf{P} = \mathbf{I}_c$ but $\mathbf{PP}'$ may not be an identity matrix.

Square matrices having orthonormal rows are in a special class: their columns are also orthonormal (why?). The matrix $\boldsymbol{P}$, say, is then such that

$$\mathbf{PP}' = \mathbf{P}'\mathbf{P} = \boldsymbol{I}. \tag{5.25}$$

This equation defines $\mathbf{P}$ as being an *orthogonal matrix*; it implies that $\boldsymbol{P}$ is square and that $\boldsymbol{P}$ has orthonormal rows and orthonormal columns. It can also be seen that $\boldsymbol{P}$ is nonsingular since the absolute value of its determinant is equal to 1. Furthermore, the inverse of $\boldsymbol{P}$ is equal to $\boldsymbol{P}'$. These are characteristics of any orthogonal matrix $\boldsymbol{P}$. Actually, it can be shown that any two of the conditions (i) $\boldsymbol{P}$ is a square matrix, (ii) $\mathbf{P}'\mathbf{P} = \mathbf{I}$, and (iii) $\mathbf{PP}' = \mathbf{I}$ imply the third, that is, imply that $\boldsymbol{P}$ is orthogonal (see Exercise 5.20).

A simple property of orthogonal matrices is that products of them are orthogonal. (see Exercise 5.9).

Theorem 5.2 (The QR Decomposition) *Let* $\boldsymbol{A}$ *be an* $n \times p$ *matrix of rank* p $(n \geq p)$. *Then,* $\boldsymbol{A}$ *can be written as* $\boldsymbol{A} = \boldsymbol{QR}$, *where* $\boldsymbol{Q}$ *is an* $n \times p$ *matrix whose columns are orthonormal and* $\boldsymbol{R}$ *is an upper-triangular matrix of order* $p \times p$.

This theorem is based on applying the so-called *Gram-Schmidt* orthogonalization of the columns of $\boldsymbol{A}$. Details of the proof can be found in Harville(1997, Section 6.4).

Example 5.10

$$\mathbf{A} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & -\sqrt{3} & 0 \\ 1 & 1 & -2 \end{bmatrix}$$

is an orthogonal matrix, as the reader may easily verify.

The boundless variety of matrices that are orthogonal includes many that are carefully proscribed, three of which are now briefly illustrated and described.

5.4.1 Special Cases

Helmert, Givens, and Householder matrices are all orthogonal matrices.

Helmert Matrices The Helmert matrix of order 4×4 is

$$\mathbf{H}_4 = \begin{bmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix}.$$

The Helmert matrix $\mathbf{H}_n$ of order $n \times n$ has $n^{-\frac{1}{2}}\mathbf{1}'_n$ for its first row, and each of its other $n-1$ rows for $i = 1, \dots, n-1$, has the partitioned form

$$\left[\mathbf{1}'_\imath \,|\, -i \,|\, 0\mathbf{1}'_{n-\imath-1}\right] / \sqrt{\lambda_i} \quad \text{with } \lambda_\imath = i(i+1).$$

The notation $0\mathbf{1}'$ used here simply indicates a row vector with every element zero.

Givens Matrices The orthogonal matrix

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

is a Givens matrix of order 2×2. It is the basis of Givens matrices of order higher than 2×2. Those of order 3×3 are

$$\mathbf{G}_{12} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}_{13} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\mathbf{G}_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma \\ 0 & -\sin\gamma & \cos\gamma \end{bmatrix}. \tag{5.26}$$

The general form of $\mathbf{G}_{rs} = \mathbf{G}_{sr}$ of order $n \times n$ is an identity matrix except for four elements: $g_{rr} = g_{ss} = \cos\theta_{rs}$ and, for $r > s$, $-g_{rs} = g_{sr} = \sin\theta_{rs}$. All such matrices are orthogonal, as are products of any number of them.

A sometimes useful operation is that of triangularizing a square matrix; that is, if $\mathbf{A}$ is square, of premultiplying it by some matrix $\mathbf{G}$ so that $\mathbf{GA}$ is an upper triangular matrix. This can be done by deriving $\mathbf{G}$ as a product of Givens matrices. The operation of going from $\mathbf{A}$ to $\mathbf{GA}$ is called a *Givens transformation*. A lower triangular matrix can be obtained from $\mathbf{A}$ in similar fashion, by postmultiplying $\mathbf{A}$ by Givens matrices.

Householder Matrices Other orthogonal matrices useful for triangularizing square matrices are the Householder matrices, the general form of which is

$$\mathbf{H} = \mathbf{I} - 2\mathbf{hh}' \quad \text{for} \quad \mathbf{h}'\mathbf{h} = 1, \tag{5.27}$$

with $\mathbf{h}$, obviously, being a non-null column vector. Then $\mathbf{H}$ is not only orthogonal but also symmetric; and for any non-null vector

$$\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \quad \text{there exists} \quad \mathbf{h}' = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix}$$

such that

$$\mathbf{Hx} = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{for} \quad \begin{cases} \lambda = -(\text{sign of } x_1)\sqrt{\mathbf{x}'\mathbf{x}} \\ h_1 = \sqrt{\frac{1}{2}(1 - x_1/\lambda)} \\ h_i = -x_i/2h_1\lambda \quad \text{for} \quad i = 2, 3, \ldots, n. \end{cases}$$

Example 5.11 *For* $\mathbf{x}' = [1 \quad 2 \quad 3 \quad 5 \quad 5]$ *we find* $\mathbf{h}' = \frac{1}{12}[9 \quad 2 \quad 3 \quad 5 \quad 5]$ *and* $\mathbf{Hx} = [-8 \quad 0 \quad 0 \quad 0 \quad 0]'$.

Triangularization of a square matrix $\mathbf{A}$ proceeds by developing Householder matrices from successive columns of $\mathbf{A}$.

5.5 PARAMETERIZATION OF ORTHOGONAL MATRICES

The n^2 elements of an orthogonal matrix $\boldsymbol{A}$ of order $n \times n$ are not independent. This follows from the fact that $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_n$ which implies that the elements of $\boldsymbol{A}$ are subject to $n(n+1)/2$ equality constraints. Hence, the number of independent elements is $n^2 - n(n+1)/2 = n(n-1)/2$. This means that $\boldsymbol{A}$ can be represented by $n(n-1)/2$ independent parameters. Knowledge of such a representation is needed in order to generate orthogonal matrices that are used in several statistical applications. Khuri and Myers (1981) adopted this approach to construct response surface designs that are robust to non-normality of the distribution of the error in the response surface model. Another application is the generation of random orthogonal matrices to be used in simulation experiments [see Heiberger et al. (1983) and Anderson et al. (1987)].

Khuri and Good (1989) reviewed several methods to parameterize an orthogonal matrix. Two of these methods are described here.

I. **Exponential Representation** [Gantmacher (1959)]
If $\boldsymbol{A}$ is an orthogonal matrix with determinant equal to 1, then it can be represented as

$$\boldsymbol{A} = \exp(\boldsymbol{T}), \tag{5.28}$$

where $\boldsymbol{T}$ is a skew-symmetric matrix. The elements of $\boldsymbol{T}$ above its main diagonal can be used to parameterize $\boldsymbol{A}$. The exponential function in (5.28) is represented as the sum of the infinite series of matrices,

$$\exp(\boldsymbol{T}) = \sum_{i=0}^{\infty} \frac{1}{i!}\boldsymbol{T}^i, \tag{5.29}$$

where $\boldsymbol{T}^0 = \boldsymbol{I}_n$.

II. **Cayley's Representation** [Gantmacher (1959, p. 289)]
If $\boldsymbol{A}$ is an orthogonal matrix of order $n \times n$ that does not have the eigenvalue -1, then it can be written in Cayley's form, namely,

$$\boldsymbol{A} = (\boldsymbol{I}_n - \boldsymbol{U})(\boldsymbol{I}_n + \boldsymbol{U})^{-1},$$

where $\boldsymbol{U}$ is a skew-symmetric matrix of order $n \times n$.

5.6 QUADRATIC FORMS

The illustration (5.19),

$$\sum_{i=1}^{n} (x_i - \bar{x}.)^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

is the product of a row vector $\mathbf{x}'$, a matrix $\mathbf{C}$, and the column vector $\mathbf{x}$. It is called a *quadratic form*; its general form for any matrix $\mathbf{A}$ is $\mathbf{x}'\mathbf{A}\mathbf{x}$. Expressions of this form have many uses, particularly in the general theory of analysis of variance in statistics wherein, with appropriate choice of $\mathbf{A}$, any sum of squares can be represented as $\mathbf{x}'\mathbf{A}\mathbf{x}$.

Consider the example

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [\, x_1 \quad x_2 \quad x_3 \,] \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Straightforward multiplication gives

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 &+ 4x_2x_1 + 2x_3x_1 + 2x_1x_2 \\ &+ 7x_2^2 - 2x_3x_2 + 3x_1x_3 + 6x_2x_3 + 5x_3^2, \end{aligned} \tag{5.30}$$

which simplifies to $\mathbf{x}'\mathbf{A}\mathbf{x} =$

$$\begin{aligned} x_1^2 + x_1x_2(4+2) + x_1x_3(2+3) + 7x_2^2 + x_2x_3(-2+6) + 5x_3^2 \\ = x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3. \end{aligned} \tag{5.31}$$

This is a quadratic function of the x's; hence the name quadratic form. Two characteristics of its development are noteworthy. First, in (5.30) we see that $\mathbf{x}'\mathbf{A}\mathbf{x}$ is the sum of products of all possible pairs of the x_i's, each multiplied by an element of $\mathbf{A}$; thus in (5.30) the second term, $4x_2x_1$, is x_2x_1 multiplied by the element of $\mathbf{A}$ in the second row and first column. Second, in simplifying (5.30) to (5.31) we see that the coefficient of x_1x_2, for example, is the sum of two elements in $\mathbf{A}$: the one in the first column and second row plus that in the second column and first row. These results are true generally.

If $\boldsymbol{x}$ is a vector of order $n \times 1$ with elements x_i for $i = 1, 2, \ldots, n$, and if $\mathbf{A}$ is a square matrix of order $n \times n$ with elements a_{ij} for $i, j = 1, 2, \ldots, n$, then

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= \sum_i \sum_j x_i x_j a_{ij} \quad \text{[similar to (5.30)]}\\ &= \sum_i x_i^2 a_{ii} + \sum_{i \neq j}\sum x_i x_j a_{ij}\end{aligned}$$

and as in (5.31) this is

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + \sum_{j>i}\sum x_i x_j \left(a_{ij} + a_{ji}\right). \tag{5.32}$$

Thus $\mathbf{x}'\mathbf{A}\mathbf{x}$ is the sum of squares of the elements of $\mathbf{x}$, each square multiplied by the corresponding diagonal element of $\mathbf{A}$, plus the sum of products of the elements of $\mathbf{x}$, each product multiplied by the sum of the corresponding elements of $\mathbf{A}$; that is, the product of the ith and jth element of $\mathbf{x}$ is multiplied by $(a_{ij} + a_{ji})$.

Returning to the example, note that

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3\\ &= x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2\,(1+5) + x_1x_3\,(1+4) + x_2x_3\,(0+4)\\ &= \mathbf{x}'\begin{bmatrix} 1 & 1 & 1\\ 5 & 7 & 0\\ 4 & 4 & 5\end{bmatrix}\mathbf{x}.\end{aligned}$$

In this way we see that

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 2 & 3\\ 4 & 7 & 6\\ 2 & -2 & 5\end{bmatrix}\mathbf{x} \text{ is the same as } \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 1 & 1\\ 5 & 7 & 0\\ 4 & 4 & 5\end{bmatrix}\mathbf{x},$$

where $\mathbf{B}$ is different from $\mathbf{A}$. Note that the quadratic form is the same, even though the associated matrix is not the same. In fact, there is no unique matrix $\mathbf{A}$ for which any particular quadratic form can be expressed as $\mathbf{x}'\mathbf{A}\mathbf{x}$. Many matrices can be so used. Each one has the same diagonal elements, and in each of them the sum of each pair of symmetrically placed off-diagonal elements a_{ij} and a_{ji} must be the same; for example, (5.30) can also be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 2342 & -789\\ -2336 & 7 & 1.37\\ 794 & 2.63 & 5\end{bmatrix}\mathbf{x}. \tag{5.33}$$

In particular, if we write

$$\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2\,(3+3) + x_1x_3\left(2\tfrac{1}{2} + 2\tfrac{1}{2}\right) + x_2x_3\,(2+2)$$

we see that it can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 3 & 2\frac{1}{2} \\ 3 & 7 & 2 \\ 2\frac{1}{2} & 2 & 5 \end{bmatrix}\mathbf{x} \tag{5.34}$$

where $\mathbf{A}$ is now a symmetric matrix. As such it is unique; that is to say, for any particular quadratic form there is a unique symmetric matrix $\mathbf{A}$ for which the quadratic form can be expressed as $\mathbf{x}'\mathbf{A}\mathbf{x}$. It can be found in any particular case by rewriting the quadratic $\mathbf{x}'\mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is not symmetric as $\mathbf{x}'[\frac{1}{2}(\mathbf{A}+\mathbf{A}')]\mathbf{x}$, because $\frac{1}{2}(\mathbf{A}+\mathbf{A}')$ is symmetric. For example, if $\mathbf{A}$ is the matrix used in (5.33), it is easily observed that $\frac{1}{2}(\mathbf{A}+\mathbf{A}')$ is the symmetric matrix used in (5.34).

Taking $\mathbf{A}$ as symmetric with $a_{ij} = a_{ji}$, we see from (5.32) that the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + 2\sum_{j>i}\sum x_i x_j a_{ij}. \tag{5.35}$$

For example,

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \mathbf{x}'\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}\mathbf{x} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3). \end{aligned}$$

The importance of the symmetric $\mathbf{A}$ is that, when writing any particular quadratic function as $\mathbf{x}'\mathbf{A}\mathbf{x}$, there are many different matrices that can be used, but there is only one symmetric matrix—and it is unique for that particular function. Therefore, when dealing with quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$, we always take $\mathbf{A}$ as symmetric. This is convenient not only because the symmetric $\mathbf{A}$ is unique for any particular quadratic form, but also because symmetric matrices have many properties that are useful in studying quadratic forms, particularly those associated with analysis of variance. Hereafter, whenever we deal with a quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$, we assume $\mathbf{A} = \mathbf{A}'$.

A slightly more general (but not so useful) function is the second-degree function in two sets of variables $\mathbf{x}$ and $\mathbf{y}$, say. For example,

$$\begin{aligned} \mathbf{x}'\mathbf{M}\mathbf{y} &= [x_1 \quad x_2]\begin{bmatrix} 2 & 4 & 3 \\ 7 & 6 & 5 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= 2x_1y_1 + 4x_1y_2 + 3x_1y_3 + 7x_2y_1 + 6x_2y_2 + 5x_2y_3. \end{aligned}$$

It is called a *bilinear form* and, as illustrated here, its matrix $\mathbf{M}$ does not have to be square as does the matrix in a quadratic form. Clearly, quadratic forms are special cases of bilinear forms—when $\mathbf{M}$ is square and $\mathbf{y} = \mathbf{x}$.

5.7 POSITIVE DEFINITE MATRICES

All quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$ are zero for $\mathbf{x} = \mathbf{0}$. For some matrices $\mathbf{A}$ the corresponding quadratic form is zero *only* for $\mathbf{x} = \mathbf{0}$. For example,

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2,\end{aligned}$$

and, by the nature of this last expression, we see that (for elements of $\mathbf{x}$ being real numbers) it is positive unless all elements of $\mathbf{x}$ are zero, that is, $\mathbf{x} = \mathbf{0}$. Such a quadratic form is described as being positive definite. More formally,

$$\text{when } \mathbf{x}'\mathbf{A}\mathbf{x} > 0 \text{ for all } \mathbf{x} \text{ other than } \mathbf{x} = \mathbf{0},$$

then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive definite* quadratic form, and $\mathbf{A} = \mathbf{A}'$ is correspondingly a *positive definite* (p.d.) matrix.

There are also symmetric matrices $\mathbf{A}$ for which $\mathbf{x}'\mathbf{A}\mathbf{x}$ is zero for some non-null $\mathbf{x}$ as well as for $\mathbf{x} = \mathbf{0}$; for example,

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_1x_3 - 6x_2x_3 \\ &= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2.\end{aligned}$$

This is zero for $\mathbf{x}' = [2 \quad 1 \quad 3]$, and for any scalar multiple thereof, as well as for $\mathbf{x} = \mathbf{0}$. This kind of quadratic form is called positive semidefinite and has for its formal definition:

$$\text{when } \mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \textit{ and } \mathbf{x}'\mathbf{A}\mathbf{x} = 0 \text{ for some } \mathbf{x} \neq \mathbf{0}$$

then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive semidefinite* quadratic form and hence $\mathbf{A} = \mathbf{A}'$ is a *positive semidefinite* (p.s.d.) matrix. The two classes of (forms and) matrices taken together, positive definite and positive semidefinite, are called *non-negative definite* (n.n.d.).

Example 5.12 *The sum of squares of (5.19),*

$$\sum_{\iota=1}^{n}(x_\iota - \bar{x})^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

is a positive semidefinite quadratic form because it is positive, except for being zero when all the x_ι's are equal. Its matrix, $\mathbf{I} - \bar{\mathbf{J}}$, which is idempotent, is also p.s.d., as are all symmetric idempotent matrices (except $\mathbf{I}$, which is the only p.d. idempotent matrix).

Unfortunately there is no universal agreement on the definition of positive semidefinite. Most writers use it with the meaning defined here, but some use it in the sense of meaning non-negative definite. But on one convention there is universal agreement: p.d., p.s.d., and n.n.d. matrices are always taken as being symmetric. This is so because the definitions of these matrices are in terms of quadratic forms which can always be expressed utilizing symmetric matrices.

Theorem 5.3

(a) *For a matrix* $\mathbf{X}$, *the product* $\mathbf{X}'\mathbf{X}$ *is positive definite if* $\mathbf{X}$ *is of full column rank and is positive semidefinite otherwise.*

(b) *For a matrix* $\mathbf{Y}$, *the product* $\mathbf{Y}\mathbf{Y}'$ *is positive definite if* $\mathbf{Y}$ *is of full row rank and is positive semidefinite otherwise.*

Proof. See Exercise 5.23. ■

REFERENCES

Andreson, T. W., Olkin, I., and Underhill, L. G. (1987). Generation of random orthogonal matrices. *SIAM J. Sci. Stat. Comput.*, 8, 625–629.

Gantmacher, F. R. (1959). *The Theory of Matrices*, Vols. I and II. Chelsea, New York.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

Heiberger, R. M., Velleman, P. F., and Ypelaar, M. A. (1983). Generating test data with independently controllable features for multivariate general linear forms. *J. Am. Stat. Assoc.*, 78, 585–595.

Khuri, A. I. and Good, I. J. (1989). The parameterization of orthogonal matrices: A review mainly for statisticians. *South African Stat. J.*, 23, 231–250.

Khuri, A. I. and Myers, R. H. (1981). Design related robustness of tests in regression models. *Comm. Stat. Theory Methods*, 10, 223–235.

EXERCISES

5.1 Show that $\begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}$ is symmetric.

5.2 If $\mathbf{x}$ and $\mathbf{y}$ are $n \times 1$ column vectors and $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, which of the following expressions are undefined? Of those that are defined, which are bilinear forms, quadratic forms?

(a) $\mathbf{y} = \mathbf{A}\mathbf{x}$.

(b) $\mathbf{x}\mathbf{y} = \mathbf{A}'\mathbf{B}$.

(c) $\mathbf{x}'\mathbf{B}\mathbf{x}$.

(d) $\mathbf{y}\mathbf{B}\mathbf{x}$.

(e) $\mathbf{y}'\mathbf{B}'\mathbf{A}\mathbf{x}$.

(f) $\mathbf{x}' = \mathbf{y}'\mathbf{B}'$.

(g) $\mathbf{x}'\mathbf{A}\mathbf{y}$.

(h) $\mathbf{y}'\mathbf{A}'\mathbf{B}\mathbf{y}$.

(i) $\mathbf{x}\mathbf{y}' = \mathbf{B}'$.

5.3 If $\mathbf{A}$ is skew symmetric, prove that (a) $a_{ii} = 0$ and $a_{ij} = -a_{ji}$; and (b) $\mathbf{I} + \mathbf{A}$ is positive definite.

5.4 For a square matrix $\mathbf{A}$, prove that it is the sum of a symmetric and a skew-symmetric matrix.

5.5 If the product of two symmetric matrices is symmetric, prove that the matrices commute in multiplication.

5.6 Show that if $\mathbf{X}'\mathbf{X} = \mathbf{X}$, then $\mathbf{X} = \mathbf{X}' = \mathbf{X}^2$

5.7 (a) Show that the only real symmetric matrix whose square is null is the null matrix itself.

(b) Explain why $\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$ implies $\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}$. (**Note:** The matrix $\boldsymbol{G}$ is a g-inverse of $\boldsymbol{X}'\boldsymbol{X}$. Such an inverse will be defined in Chapter 8).

5.8 If $X_{r\times c} = [\boldsymbol{x}_1\, \boldsymbol{x}_2\, \ldots\, \boldsymbol{x}_c]$, prove that for symmetric $\mathbf{A}$ $tr[(\mathbf{A}\mathbf{X}\mathbf{X}')^2] = \Sigma_{j=1}^{c}\Sigma_{k=1}^{c}(\mathbf{x}_j'\,\mathbf{A}\mathbf{x}_k)^2$.

5.9 Prove that a product of orthogonal matrices is orthogonal.

5.10 (a) If $\mathbf{A}$ is idempotent and symmetric, prove that it is positive semidefinite.

(b) When $\mathbf{X}$ and $\mathbf{Y}$ are idempotent, prove that $\mathbf{X}\mathbf{Y}$ is, provided $\mathbf{X}$ and $\mathbf{Y}$ commute in multiplication.

(c) Prove that $\mathbf{I} + \mathbf{K}\mathbf{K}'$ is positive definite for real $\mathbf{K}$.

5.11 Using $\mathbf{x}' = (1, 3, 5, 7, 9)$, derive or state the numerical value of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ such that

(a) $1^2 + 3^2 + 5^2 + 7^2 + 9^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$;

(b) $(1 + 3 + 5 + 7 + 9)^2/5 = \mathbf{x}'\mathbf{B}\mathbf{x}$;

(c) $(1 - 5)^2 + (3 - 5)^2 + (5 - 5)^2 + (7 - 5)^2 + (9 - 5)^2 = \mathbf{x}'\boldsymbol{C}\boldsymbol{x}$.

5.12 Explain why

(a) $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{x}$, even when $\mathbf{A}$ is not symmetric;

(b) $\mathbf{x}'\mathbf{B}\mathbf{x} = tr(\mathbf{x}'\mathbf{B}\mathbf{x})$;

(c) $\mathbf{x}'\mathbf{C}\mathbf{x} = tr(\mathbf{C}\mathbf{x}\mathbf{x}')$.

5.13 (a) Suppose that the columns of the matrix $\boldsymbol{P}_{r\times c}$ are orthonormal, that is, $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_c$. Give an example to demonstrate that the rows of $\boldsymbol{P}$ are not necessarily orthonormal, that is, $\boldsymbol{P}\boldsymbol{P}'$ is not necessarily an identity matrix.

(b) Show that a square matrix having orthonormal rows must also have orthonormal columns.

5.14 Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric idempotent matrices.

(a) If $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}$, show that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

(b) Show that the reverse of (a) is true, that is, if $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite, then $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}$.

(c) Show that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}$ if and only if $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{B}$.

5.15 Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of order $m \times n$. Show that

$$[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \leq tr(\boldsymbol{A}'\boldsymbol{A})tr(\boldsymbol{B}'\boldsymbol{B}).$$

This is known as the *Cauchy–Schwarz inequality* for matrices. Equality holds if and only if one of the matrices is a multiple of the other.

5.16 Suppose that $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$ and let $\boldsymbol{B}$ be a skew-symmetric matrix of order $n \times n$. Show that $tr(\boldsymbol{AB}) = 0$.

5.17 Let $\boldsymbol{A}$ be a matrix of order $n \times n$. Suppose that $tr(\boldsymbol{PA}) = 0$ for every skew-symmetric matrix of order $n \times n$. Show that $\boldsymbol{A}$ is symmetric.

5.18 Suppose that the matrix $\boldsymbol{A}$ is positive definite and that the matrix $\boldsymbol{P}$ is nonsingular. Show that $\boldsymbol{P}'\boldsymbol{AP}$ is positive definite. What can be said about $\boldsymbol{P}'\boldsymbol{AP}$ if $\boldsymbol{A}$ is assumed to be positive semidefinite?

5.19 Let $\boldsymbol{V}$ be an $n \times n$ symmetric positive definite matrix such that $\boldsymbol{V} = \boldsymbol{P}'\boldsymbol{P}$, where $\boldsymbol{P}$ is a matrix of order $n \times n$, and let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix. Show that

(a) $\boldsymbol{P}$ is a nonsingular matrix.

(b) $\boldsymbol{PAP}'$ is idempotent if and only if $\boldsymbol{AV}$ is idempotent.

(c) If $\boldsymbol{AV}$ is idempotent, then $r(\boldsymbol{PAP}') = tr(\boldsymbol{AV})$.

5.20 Show that any two of the conditions,
(i) $\boldsymbol{P}$ is a square matrix of order $n \times n$, (ii) $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_n$, and (iii) $\boldsymbol{PP}' = \boldsymbol{I}_n$ imply the third, that is, imply that $\boldsymbol{P}$ is orthogonal.

5.21 Consider the following two-way crossed classification model concerning two factors, denoted by A and B, having 4 and 3 levels, respectively, with two replications per treatment combination:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

where α_i represents the ith level of factor A, β_j represents the jth level of factor B, $(\alpha\beta)_{ij}$ is the corresponding interaction term, and ϵ_{ijk} is an experimental error term, $i = 1, 2, 3, 4$; $j = 1, 2, 3$; $k = 1, 2$. Consider the following sums of squares from the corresponding analysis of variance table:

$$SS_A = 6\sum_{i=1}^{4}(\bar{y}_{i..} - \bar{y}_{...})^2$$

$$SS_B = 8\sum_{j=1}^{3}(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$SS_{AB} = 2\sum_{i=1}^{4}\sum_{j=1}^{3}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2,$$

where

$$
\begin{aligned}
\bar{y}_{i..} &= \frac{1}{6}\sum_{j=1}^{3}\sum_{k=1}^{2} y_{ijk} \\
\bar{y}_{.j.} &= \frac{1}{8}\sum_{i=1}^{4}\sum_{k=1}^{2} y_{ijk} \\
\bar{y}_{ij.} &= \frac{1}{2}\sum_{k=1}^{2} y_{ijk} \\
\bar{y}_{...} &= \frac{1}{24}\sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{k=1}^{2} y_{ijk}.
\end{aligned}
$$

Derive the matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, and $\boldsymbol{P}_3$ such that

$$
\begin{aligned}
SS_A &= \boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y} \\
SS_B &= \boldsymbol{y}'\boldsymbol{P}_2\boldsymbol{y} \\
SS_{AB} &= \boldsymbol{y}'\boldsymbol{P}_3\boldsymbol{y},
\end{aligned}
$$

where $\boldsymbol{y}$ is the column vector of all 24 values of y_{ijk} with k varying first from 1 to 2, followed by j varying from 1 to 3, and then i varying from 1 to 4.

5.22 Let $\boldsymbol{A}$ be an $n \times n$ positive definite matrix and let $\boldsymbol{d}$ be a vector of n elements. Consider the matrix,

$$\boldsymbol{B} = \boldsymbol{A} - \frac{1}{c}\boldsymbol{d}\boldsymbol{d}',$$

where c is a scalar such that $c > \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d}$. Show that the inverse of $\boldsymbol{B}$ exists and is given by

$$\boldsymbol{B}^{-1} = \boldsymbol{A}^{-1} + \gamma\boldsymbol{A}^{-1}\boldsymbol{d}\boldsymbol{d}'\boldsymbol{A}^{-1},$$

where $\gamma = (c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})^{-1}$.

5.23 Show that if

(a) $\boldsymbol{X}$ is of full column rank, then $\boldsymbol{X}'\boldsymbol{X}$ is positive definite, otherwise $\boldsymbol{X}'\boldsymbol{X}$ is positive semidefinite.

(b) $\boldsymbol{Y}$ is of full row rank, then $\boldsymbol{Y}\boldsymbol{Y}'$ is positive definite, otherwise $\boldsymbol{Y}\boldsymbol{Y}'$ is positive semidefinite.

6

Eigenvalues and Eigenvectors

This chapter discusses a particular linear transformation which maps a vector into a scalar multiple of itself. More specifically, let $\boldsymbol{A}$ be a matrix of order $n \times n$ which serves as a linear transformation of a vector $\boldsymbol{u}$ such that

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}, \tag{6.1}$$

where λ is a scalar. It is of interest to know the conditions under which $\boldsymbol{u}$ and λ can exist so that (6.1) is valid for $\boldsymbol{u} \neq 0$.

6.1 DERIVATION OF EIGENVALUES

Equation (6.1) can be written as

$$(\mathbf{A} - \lambda \boldsymbol{I}_n)\mathbf{u} = \mathbf{0}, \tag{6.2}$$

from which we know that if $\mathbf{A} - \lambda \boldsymbol{I}_n$ is nonsingular, the only solution is $\mathbf{u} = \mathbf{0}$. But if we were to have a non-null solution, then (6.2) indicates that the columns of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ are linearly dependent and therefore the rank of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ must be less than n, which implies that $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ is a singular matrix (see Section 4.15). This shows that the determinant of the matrix in (6.2) must be equal to zero, that is,

$$|\mathbf{A} - \lambda \boldsymbol{I}_n| = 0. \tag{6.3}$$

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

Hence, (6.3) is the condition for $\mathbf{u}$ and λ to exist such that (6.2) is true, that is, pick λ so that the determinant of $\mathbf{A} - \lambda \boldsymbol{I}_n$ is zero.

Equation (6.3) is called the *characteristic equation* of $\mathbf{A}$. For $\mathbf{A}$ of order $n \times n$, the characteristic equation is a polynomial equation in λ of degree n, with n roots to be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$, some of which may be zero. These roots are called *latent roots, characteristic roots, or eigenvalues*. Corresponding to each root λ_i is a vector $\mathbf{u}_i$ satisfying (6.1):

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for} \quad i = 1, \ldots, n, \tag{6.4}$$

and these vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are correspondingly called *latent vectors, characteristic vectors*, or *eigenvectors*. Eigenvalues and eigenvectors will be used in this book. The prefix eigen originated from the German word "eigen" which means "characteristic," "particular," "special," or "unique to."

Example 6.1 *The matrix*

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has the characteristic equation

$$\left| \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0; \quad \textit{that is,} \quad \begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0. \tag{6.5}$$

Expanding the determinant in (6.5) gives

$$(1-\lambda)^2 - 36 = 0; \quad \textit{that is,} \quad \lambda = -5 \textit{ or } 7.$$

Note that characteristic equations are, from (6.3), always of the form shown in (6.5): equated to zero is the determinant of $\mathbf{A}$ *amended by subtracting* λ *from each diagonal element.*

The derivation of eigenvectors corresponding to solutions of a characteristic equation is discussed in Section 6.3, but meanwhile it can be seen here that

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5 \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \tag{6.6}$$

these being examples of (6.1). Thus $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ *is an eigenvector corresponding to the eigenvalue –5, and* $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ *is a vector for the root 7.*

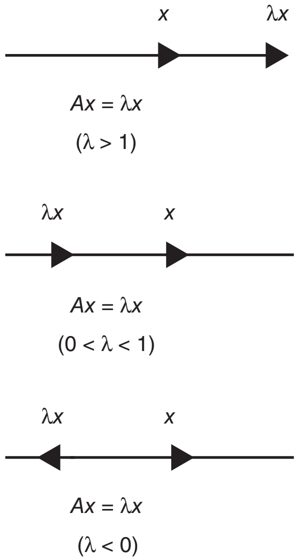


Figure 6.1 *The Matrix **A** with Its Eigenvalue λ and Eigenvector **x**.*

Example 6.2 *The characteristic equation for*

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \quad \textit{is} \quad \begin{vmatrix} 2-\lambda & 2 & 0 \\ 2 & 1-\lambda & 1 \\ -7 & 2 & -3-\lambda \end{vmatrix} = 0.$$

An expansion of the determinant gives, after simple arithmetic,

$$\lambda^3 - 13\lambda + 12 = 0, \quad \textit{equivalent to} \quad (\lambda - 1)(\lambda^2 + \lambda - 12) = 0.$$

Solutions are $\lambda = 1,\ 3$*, and –4, and these are the eigenvalues of* ***A***.

Note: If an eigenvalue λ of a matrix $\boldsymbol{A}$ is greater than one, then the action of $\boldsymbol{A}$ on a corresponding eigenvector $\boldsymbol{x}$ amounts to "stretching" it without changing its direction. However, if $0 < \lambda < 1$, then $\boldsymbol{A}$ acts on $\boldsymbol{x}$ by shrinking its length without a change in direction. In the event $\lambda < 0$, then $\boldsymbol{Ax}$ has an opposite direction to that of $\boldsymbol{x}$. These three cases are illustrated in Figure 6.1.

6.1.1 Plotting Eigenvalues

Some of the eigenvalues of a matrix can be complex numbers with each having a real part and an imaginary part. Figure 6.2 shows a scatter plot of eigenvalues of a matrix of order 50×50 whose elements were randomly generated from a standard normal distribution with mean zero and standard deviation 1. Each point in the plot represents an eigenvalue with a real part (on the horizontal axis) and a possible imaginary part (on the vertical axis). If the latter value is equal to zero, then the eigenvalue is a real number. This plot provides a

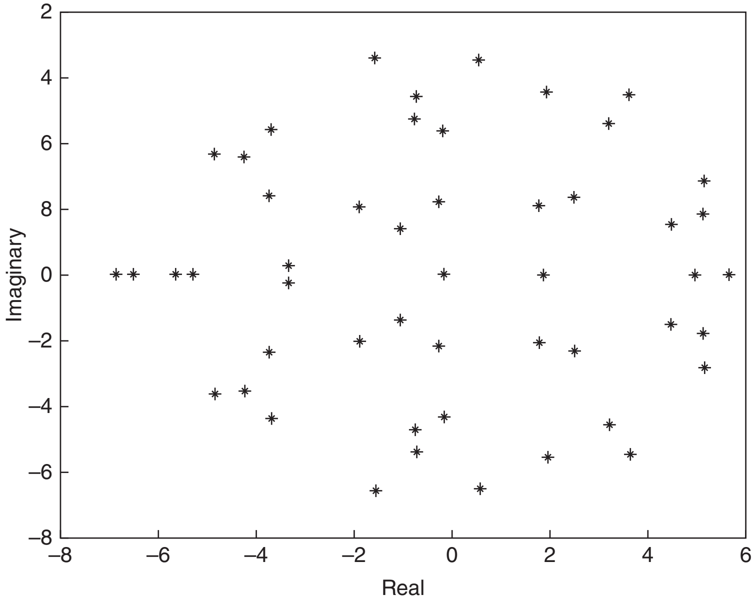


Figure 6.2 *A Scatter Plot of Eigenvalues of a Matrix of Order* 50×50.

convenient way to graphically describe the distribution of the eigenvalues of a matrix. The plot was derived using the MATLAB computer package (see Chapter 16 for more details).

6.2 ELEMENTARY PROPERTIES OF EIGENVALUES

Several properties of eigenvalues that stem directly from their definition in (6.1) and the resulting characteristic equation (6.3) command attention before considering eigenvectors.

6.2.1 Eigenvalues of Powers of a Matrix

The defining equation is $\mathbf{Au} = \lambda\mathbf{u}$ of (6.1). Premultiplying this by $\mathbf{A}$ and using (6.1) again gives

$$\mathbf{A}^2\mathbf{u} = \mathbf{A}\lambda\mathbf{u} = \lambda\mathbf{Au} = \lambda(\lambda\mathbf{u}) = \lambda^2\mathbf{u}.$$

Comparing this equation with (6.1), we see that it defines λ^2 as being an eigenvalue of $\mathbf{A}^2$. Similarly, λ^3 is an eigenvalue of $\mathbf{A}^3 : \mathbf{A}^3\mathbf{u} = \mathbf{A}\lambda^2\mathbf{u} = \lambda^2\mathbf{Au} = \lambda^3\mathbf{u}$, and in general λ^k is an eigenvalue of $\mathbf{A}^k$:

$$\mathbf{A}^k\mathbf{u} = \lambda^k\mathbf{u}. \tag{6.7}$$

Furthermore, when $\mathbf{A}$ is nonsingular (6.1) gives $\mathbf{u} = \mathbf{A}^{-1}\lambda\mathbf{u} = \lambda\mathbf{A}^{-1}\mathbf{u}$ and so

$$\mathbf{A}^{-1}\mathbf{u} = \lambda^{-1}\mathbf{u}. \tag{6.8}$$

Hence, when λ is an eigenvalue of $\boldsymbol{A}$, then λ^k is an eigenvalue of $\mathbf{A}^k$ where k is positive if $\mathbf{A}$ is singular, and k is positive or negative if $\mathbf{A}$ is nonsingular. In particular, when $\mathbf{A}$ is nonsingular with eigenvalue λ, the inverse $\mathbf{A}^{-1}$ has $1/\lambda$ as an eigenvalue. (Recall that $k = 0$ gives $\mathbf{A}^k = \mathbf{A}^0 = \mathbf{I}$).

6.2.2 Eigenvalues of a Scalar-by-Matrix Product

Multiplying $\mathbf{Au} = \lambda\mathbf{u}$ of (6.1) by a scalar c gives

$$c\mathbf{Au} = c\lambda\mathbf{u}. \tag{6.9}$$

Hence, when λ is an eigenvalue of $\boldsymbol{A}$, we have $c\lambda$ being an eigenvalue of $c\mathbf{A}$.

Rewriting (6.9) as $\mathbf{A}(c\mathbf{u}) = \lambda(c\mathbf{u})$ shows that when $\mathbf{u}$ is an eigenvector of $\boldsymbol{A}$ so also is $c\,\boldsymbol{u}$, both vectors corresponding to the same eigenvalue λ.

When $\boldsymbol{A}$ has eigenvalue λ then $\boldsymbol{A} + c\mathbf{I}$ for scalar c has eigenvalue $\lambda + c$. This is so because

$$(\mathbf{A} + c\mathbf{I})\,\mathbf{u} = \mathbf{Au} + c\mathbf{u} = \lambda\mathbf{u} + c\mathbf{u} = (\lambda + c)\,\mathbf{u}.$$

Combining this with (6.8) shows that when $(\mathbf{A} + c\mathbf{I})^{-1}$ exists, it has $1/(\lambda + c)$ as an eigenvalue.

6.2.3 Eigenvalues of Polynomials

A consequence of the two preceding sections is that when $\mathbf{A}$ has an eigenvalue λ then a polynomial in $\boldsymbol{A}$, say $f(\boldsymbol{A})$, has an eigenvalue $f(\lambda)$. For example, consider the polynomial $f(\mathbf{A}) = \mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I}$. If λ and $\boldsymbol{u}$ are a corresponding eigenvalue and vector of $\boldsymbol{A}$,

$$\mathbf{A}^3\mathbf{u} = \lambda^3\mathbf{u}, \quad \mathbf{A}^2\mathbf{u} = \lambda^2\mathbf{u}, \quad \mathbf{Au} = \lambda\mathbf{u}, \quad \text{and} \quad 5\mathbf{Iu} = 5\mathbf{u}.$$

Hence,

$$\begin{aligned}(\mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I})\mathbf{u} &= \mathbf{A}^3\mathbf{u} + 7\mathbf{A}^2\mathbf{u} + \mathbf{Au} + 5\mathbf{Iu} \\ &= \lambda^3\mathbf{u} + 7\lambda^2\mathbf{u} + \lambda\mathbf{u} + 5\mathbf{u} \\ &= \left(\lambda^3 + 7\lambda^2 + \lambda + 5\right)\mathbf{u}.\end{aligned}$$

Thus $f(\lambda) = \lambda^3 + 7\lambda^2 + \lambda + 5$ is an eigenvalue. Extension to any polynomial $f(\mathbf{A})$ gives the result that $f(\mathbf{A})$ has eigenvalue $f(\lambda)$. An important special case is that $e^{\mathbf{A}} = \sum_{i=0}^{\infty} \mathbf{A}^i/i!$ has eigenvalue e^{λ}.

Example 6.3 (continuation of Example 6.1)

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has eigenvalues –5 and 7. The eigenvalues of

$$\begin{aligned} f(\mathbf{A}) &= 2\mathbf{A}^2 + 2\mathbf{A} - 12\mathbf{I} \\ &= 2\begin{bmatrix} 37 & 8 \\ 18 & 37 \end{bmatrix} + 2\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - 12\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 64 & 24 \\ 54 & 64 \end{bmatrix} \end{aligned}$$

are given by $\begin{vmatrix} 64-\lambda & 24 \\ 54 & 64-\lambda \end{vmatrix} = 0$, *that is,* $\lambda^2 - 128\lambda + 2800 = 0$, *so that* λ *is 100 or 28. And* $100 = 2(7)^2 + 2(7) - 12 = f(7)$, *and* $28 = f(-5)$.

6.2.4 The Sum and Product of Eigenvalues

The characteristic equation (6.3) of degree 3,

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} = 0, \tag{6.10}$$

reduces to

$$-\lambda^3 + (-\lambda)^2\, tr_1(\mathbf{A}) + (-\lambda)\, tr_2(\mathbf{A}) + |\mathbf{A}| = 0 \tag{6.11}$$

using diagonal expansion of determinants and the $\mathrm{tr}_i(\mathbf{A})$ notation defined in Section 3.6. For $\boldsymbol{A}$ of order $n \times n$ this takes the form

$$\begin{aligned} &(-\lambda)^n + (-\lambda)^{n-1}\, tr_1(\mathbf{A}) + (-\lambda)^{n-2}\, tr_2(\mathbf{A}) \\ &\qquad + \cdots + (-\lambda)\, tr_{n-1}(\mathbf{A}) + tr_n(\mathbf{A}) = 0. \end{aligned} \tag{6.12}$$

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are roots of this equation, then it is equivalent to

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) = 0,$$

which expands to

$$(-\lambda)^n + (-\lambda)^{n-1}\sum \lambda_i + (-\lambda)^{n-2}\sum_{i\neq j} \lambda_i\lambda_j + \cdots + \prod_{i=1}^{n} \lambda_i = 0. \tag{6.13}$$

Equating the coefficients of $(-\lambda)^{n-1}$ in (6.12) and (6.13), and also equating the final terms, gives

$$\sum_{i=1}^{n} \lambda_i = tr_1(\mathbf{A}) = tr(\mathbf{A}) \quad \text{and} \quad \prod_{i=1}^{n} \lambda_i = tr_n(\mathbf{A}) = |\mathbf{A}|. \tag{6.14}$$

Hence, the sum of the eigenvalues of a matrix equals its trace, and their product equals its determinant.

Example 6.4 (continuation of Example 6.2)

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \quad \textit{has determinant} \quad \begin{vmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{vmatrix} = -12$$

*and its eigenvalues are 1, 3, and – 4. Their sum is 1 + 3 – 4 = 0 as is tr(**A**) = 2 + 1 – 3 = 0; and their product is* $1(3)(-4) = -12, = |\mathbf{A}|$.

6.3 CALCULATING EIGENVECTORS

Suppose λ_k is an eigenvalue of the $n \times n$ matrix $\mathbf{A}$. It is a solution of the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Calculating an eigenvector corresponding to λ_k requires finding a non-null $\mathbf{u}$ to satisfy $\mathbf{Au} = \lambda_k\mathbf{u}$, which is equivalent to solving

$$(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}. \tag{6.15}$$

We recall that $\mathbf{A} - \lambda_k\mathbf{I}$ is a singular matrix. Let ρ denote its rank which must be smaller than n. It follows that the orthogonal complement of the row space of $\boldsymbol{A} - \lambda_k\boldsymbol{I}$ is of dimension $n - \rho > 0$ and is therefore non-empty. This orthogonal complement forms a subspace of the n-dimensional Euclidean space. Any solution to (6.15) must belong to this subspace. Hence, (6.15) has at least one non-null solution $\boldsymbol{u}$. To find such a solution, ρ linearly independent equations are selected from (6.15), then $n - \rho$ elements of $\boldsymbol{u}$ are chosen arbitrarily and used to solve for the remaining ρ elements of $\boldsymbol{u}$. Thus, infinitely many non-null solutions to (6.15) can be found in this manner. These solutions form a subspace called the *eigenspace* corresponding to λ_k.

6.3.1 Simple Roots

Since λ_k is a solution to a polynomial equation it can be a solution more than once, in which case it is called a *multiple root*. We deal first with λ_k being a solution only once, in which case we call λ_k a *simple root*.

Whenever λ_k is a simple root, the rank ρ of $\mathbf{A} - \lambda_k\mathbf{I}$ is equal to $n - 1$ (see Section 7.1). Hence, $n - \rho = 1$. This implies that the eigenspace corresponding to λ_k is of dimension 1. Therefore, there exists only one linearly independent eigenvector corresponding to the simple root. This means that all values of $\mathbf{u}_k$ for a given λ_k are multiples of one another.

Example 6.5 (continuation of Example 6.2) *The eigenvalues of* $\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}$ *are* $\lambda_1 = 1$, $\lambda_2 = 3$ *and* $\lambda_3 = -4$. *For* $\lambda_1 = 1$ *we get,*

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{bmatrix}$$

Making the substitution in (6.15), we obtain

$$\begin{aligned} u_1 + 2u_2 &= 0, \\ 2u_1 + u_3 &= 0, \\ -7u_1 + 2u_2 - 4u_3 &= 0. \end{aligned}$$

Here, $\rho = 2$ and $n - \rho = 1$. Thus, one of u_1, u_2, u_3 can be given an arbitrary value. For example, we choose $u_1 = -2$. Using the first two of the above three equations, which are linearly independent, and then solving for u_2 and u_3, we get $u_2 = 1$ and $u_3 = 4$. Hence, an eigenvector corresponding to $\lambda_1 = 1$ is $(-2, 1, 4)'$.

Similarly, with $\lambda_2 = 3$, we get the equations,

$$\begin{aligned} -u_1 + 2u_2 &= 0 \\ 2u_1 - 2u_2 + u_3 &= 0 \\ -7u_1 + 2u_2 - 6u_3 &= 0. \end{aligned}$$

The first two equations are linearly independent. Assigning a value to u_1, for example $u_1 = -2$, and solving for u_2 and u_3, we get $u_2 = -1$, $u_3 = 2$. The resulting eigenvector is $(-2, -1, 2)$.

Finally, for $\lambda_3 = -4$ and following similar steps as before, we obtain $(1, -3, 13)'$ *as a corresponding eigenvector.*

6.3.2 Multiple Roots

We now deal with eigenvalues λ_k that are solutions of the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ more than one time. Any λ_k for which this occurs is called a *multiple eigenvalue* and the number of times that it is a solution is called its *multiplicity*. In general, we formulate $\boldsymbol{A}$ of order $n \times n$ as having s distinctly different eigenvalues $\lambda_1, \ldots, \lambda_s$ with λ_k having multiplicity m_k for $k = 1,\ 2, \ldots, s$ and, of course, $\sum_{k=1}^{s} m_k = n$. Note two features of this formulation. If zero is an eigenvalue (as it can be), it is one of the λ_k'a; and simple eigenvalues are also included, their multiplicities each being 1.

As already noted, each eigenvalue λ_k having multiplicity $m_k = 1$ leads to $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$ and hence, correspondingly just one linearly independent eigenvector. In contrast, for any multiple eigenvalue λ_k, the rank of $\mathbf{A} - \lambda_k\mathbf{I}$ must be ascertained for each value because, as shall be seen, this rank value plays an important role in subsequent developments.

Example 6.6 *The characteristic equation of*

$$\mathbf{A} = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad \textit{reduces to} \quad (\lambda - 1)\left(\lambda^2 - 1\right) = 0$$

so that the eigenvalues are $\lambda_1 = 1$ with $m_1 = 2$, and $\lambda_2 = -1$ with $m_2 = 1$. For $\lambda_1 = 1$,

$$\mathbf{A} - \lambda_1\mathrm{I} = \begin{bmatrix} -2 & -2 & -2 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad \textit{has rank } 1, \tag{6.16}$$

so that there are $3 - 1 = 2$ *linearly independent eigenvectors of* $\mathbf{A}$ *corresponding to* $\lambda_1 = 1$. *Using (6.16), the elements* u_1, u_2, u_3 *satisfy the following three equations:*

$$\begin{aligned} -2u_1 - 2u_2 - 2u_3 &= 0 \\ u_1 + u_2 + u_3 &= 0 \\ -u_1 - u_2 - u_3 &= . \end{aligned}$$

Only one equation is linearly independent. Selecting, for example, the second one, we have

$$u_1 + u_2 + u_3 = 0. \tag{6.17}$$

Two of the u_i'*s are linearly independent. We can therefore choose* u_2 *and* u_3 *to have arbitrary values and use (6.17) to solve for* u_1. *Since we need to have two linearly independent eigenvectors for the eigenvalue* λ_1, *we can assign two pairs of values for* u_2 *and* u_3 *so that, with the corresponding two* u_1 *values, we obtain two linearly independent eigenvectors. For example, choosing the two pairs* $(1, -1)$, $(1, 1)$ *for* u_2 *and* u_3, *the corresponding two values of* u_1 *(from using (6.17)) are 0, −2. This results in the two linearly independent eigenvectors* $(0, 1, -1)'$, $(-2, 1, 1)'$ *that correspond to* $\lambda_1 = 1$.

For $\lambda_2 = -1$ *we get, from using (6.15)*

$$\mathbf{A} - \lambda_2\mathbf{I} = \begin{bmatrix} 0 & -2 & -2 \\ 1 & 3 & 1 \\ -1 & -1 & 1, \end{bmatrix}$$

The rank of this matrix is 2. Hence, only one linearly independent eigenvector exists for λ_2. *Using a similar procedure as before, we find a corresponding eigenvector given by* $(2, -1, 1)'$. *All three eigenvectors thus obtained are linearly independent.*

Example 6.7 *The characteristic equation of*

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 \\ 2 & 3 & 0 \\ 0 & 4 & 5 \end{bmatrix} \quad \textit{reduces to} \quad (\lambda - 1)^2 (\lambda - 6) = 0,$$

so that eigenvalues are $\lambda_1 = 1$ *with* $m_1 = 2$, *and* $\lambda_2 = 6$ *with* $m_2 = 1$. *Then,*

$$\mathbf{A} - \lambda_1\mathbf{I} = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{bmatrix} \quad \textit{and} \quad \mathbf{A} - \lambda_2\mathbf{I} = \begin{bmatrix} -6 & 1 & 2 \\ 2 & -3 & 0 \\ 0 & 4 & -1 \end{bmatrix} \tag{6.18}$$

both have rank 2, so there is $3 - 2 = 1$ *linearly independent eigenvector corresponding to each of* λ_1 *and* λ_2. *Using the above two matrices, it can be verified that the following are corresponding eigenvectors:* $\mathbf{u}_1' = (1, -1, 1)$ *and* $\mathbf{u}_2' = (3, 2, 8)$. *Note that in this example although* $n = 3$ *there are only* $2(< n)$ *linearly independent eigenvectors. We refer to this subsequently.*

6.4 THE SIMILAR CANONICAL FORM

6.4.1 Derivation

Every eigenvalue λ_i has a corresponding eigenvector $\mathbf{u}_i$ for which

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for} \quad i = 1, 2, \ldots, n. \tag{6.19}$$

This is true for all n roots of the characteristic equation, which means that for λ_k of multiplicity m_k there are m_k equations like (6.19), each involving the same λ_k; if there are m_k linearly independent eigenvectors $\mathbf{u}_k$ corresponding to λ_k (as in Example 6.6), there will be one equation for each vector, whereas if there are not m_k linearly independent eigenvectors $\mathbf{u}_k$ (as in Example 6.7), there can still be m_k equations with some of the $\mathbf{u}_k$'s being repetitions of others. In all cases, the array of equations represented by 6.19 can be written as

$$\mathbf{A}\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}.$$

Then, on defining $\boldsymbol{U}$ as the matrix of n eigenvectors and $\boldsymbol{D}$ as the diagonal matrix of eigenvalues, namely,

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \quad \text{and} \quad \mathbf{D} = \begin{bmatrix}\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}, \tag{6.20}$$

we have

$$\mathbf{AU} = \mathbf{UD}. \tag{6.21}$$

$\boldsymbol{D}$ is known as the *canonical form under similarity* or equivalently as the *similar canonical form.*

The repetition of $\mathbf{u}_k$'s is illustrated for Example 6.7 by

$$\mathbf{U} = \begin{bmatrix}1 & 1 & 3 \\ -1 & -1 & 2 \\ 1 & 1 & 8\end{bmatrix} \quad \text{with} \quad \mathbf{D} = \begin{bmatrix}1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6\end{bmatrix}. \tag{6.22}$$

Through permitting this repetition whenever necessary, the form (6.21) can always be made to exist. However, in cases involving repeated $\mathbf{u}_k$'s it is clear that $\boldsymbol{U}$ is singular. But for the cases when $\boldsymbol{U}$ is nonsingular, that is, when all n eigenvectors are linearly independent, (6.21) can be expressed as

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \text{diag}\left\{\lambda_1,\, \lambda_2, \ldots, \lambda_n\right\}, \tag{6.23}$$

The product on the left is known as reduction to the similar canonical form; $\boldsymbol{D}$ on the right is the diagonal matrix of eigenvalues.

The existence of (6.23) depends on $\boldsymbol{U}$ being nonsingular. $\boldsymbol{AU} = \boldsymbol{UD}$ of (6.21) always exists, but rearranging it as $\mathbf{U}^{-1}\boldsymbol{AU} = \boldsymbol{D}$ of (6.23) requires $\boldsymbol{U}$ to be nonsingular. Within the context of eigenvalues and eigenvectors, the existence of nonsingular $\boldsymbol{U}$ is sometimes referred to as $\boldsymbol{A}$ *being diagonable*, since $\boldsymbol{D}$ is diagonal. An important theorem provides us with conditions for ascertaining whether $\boldsymbol{U}$ is singular or not. We refer to it as the diagonability theorem.

Theorem 6.1 (Diagonability Theorem) $\mathbf{A}_{n\times n}$, *having eigenvalues* λ_k *with multiplicity* m_k *for* $k = 1, 2, \ldots, s$ *and* $\sum_{k=1}^{s} m_k = n$, *has n eigenvectors that are linearly independent if and only if* $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ *for all* $k = 1, 2, \ldots, s$; *whereupon* $\mathbf{U}$ *of (6.20) is nonsingular and* $\mathbf{A}$ *is diagonable as* $\mathbf{U}^{-1}\boldsymbol{AU} = \boldsymbol{D}$ *of (6.23).*

Proof of this theorem is lengthy and is given in Section 7.1, together with prerequisite lemmas.

Existence of $\mathbf{U}^{-1}$ depends upon ascertaining if

$$r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k. \tag{6.24}$$

Each eigenvalue satisfying (6.24) is called a *regular* eigenvalue. When every eigenvalue is regular, $\mathbf{U}^{-1}$ exists and $\boldsymbol{A}$ is called a *regular matrix*. Whenever one or more eigenvalues are not regular, $\mathbf{U}^{-1}$ does not exist. Thus, a single violation of (6.24) ensures the nonexistence of $\mathbf{U}^{-1}$, whereupon $\mathbf{U}^{-1}\mathbf{AU}$ does not exist and $\boldsymbol{A}$ is said to be a *deficient* or *defective matrix*. Note, though, that there is no need to check (6.24) for simple eigenvalues (multiplicity $m_k = 1$) because, as already alluded to and proved in Section 7.1, (6.24) is always satisfied for such values.

Example 6.8 (continuation of Example 6.2) *Each eigenvalue has* $m_k = 1$, *we know that (6.24) is satisfied, and so* $\mathbf{U}^{-1}$ *exists. Assembling the eigenvectors into a matrix* $\boldsymbol{U}$ *we get*

$$\mathbf{U} = \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \quad \textit{and} \quad \mathbf{U}^{-1} = \frac{1}{70}\begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}. \tag{6.25}$$

With these values

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \textit{diag}\,\{1, 3, -4\} \quad \textit{and} \quad \mathbf{A} = \mathbf{UDU}^{-1}. \tag{6.26}$$

The reader should check these calculations.

Example 6.9 (continuation of Example 6.6) *Equation (6.16) has* $\lambda_1 = 1$ *with* $m_1 = 2$; *and* $r(\mathbf{A} - \lambda_1\mathbf{I}) = 1 = 3 - 2 = n - m_1$ *satisfies (6.24). The other eigenvalue is* $\lambda_2 = -1$ *with* $m_2 = 1$ *which we know satisfies (6.24); that is, (6.24) is satisfied by all* λ_k*'s. Hence,* $\mathbf{U}^{-1}$ *exists:*

$$\mathbf{U} = \begin{bmatrix} 0 & -2 & 2 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}, \quad \textit{with} \quad \mathbf{U}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

and it is easily checked that

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D} = diag\,\{1,\ 1,\ -1\} \quad \textit{and} \quad \mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}.$$

Example 6.10 (continuation of Example 6.7) *Here we have $\lambda_1 = 2$ with $m_1 = 2$ and from (6.18),*

$$r(\mathbf{A} - \lambda_1\mathbf{I}) = 2 \neq n - m_1 = 3 - 2 = 1.$$

Hence (6.24) is not satisfied, $\mathbf{U}^{-1}$ does not exist, and so $\mathbf{A}$ is not diagonable as $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}$; but of course $\boldsymbol{AU} = \boldsymbol{UD}$ exists, with $\boldsymbol{U}$ and $\mathbf{D}$ of (6.22) wherein the singularity of $\boldsymbol{U}$ is clearly apparent.

Since (6.24) is satisfied for every simple eigenvalue, the diagonability theorem is satisfied for any matrix having all n eigenvalues distinct. It is also satisfied for symmetric matrices, as discussed in Section 7.1. Further properties of the eigenvalues and vectors of symmetric matrices are given in Section 7.2.

6.4.2 Uses

Uses of the similar canonical form are many and varied. One of the most important applications is the ease provided by the diagonability of $\boldsymbol{A}$ for calculating powers of $\mathbf{A}$. This is because $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}$ implies

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1} \tag{6.27}$$

and hence $\mathbf{A}^2 = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}\mathbf{U}\mathbf{D}\mathbf{U}^{-1} = \mathbf{U}\mathbf{D}^2\mathbf{U}^{-1}$ and in general, for positive integers p,

$$\mathbf{A}^p = \mathbf{U}\mathbf{D}^p\mathbf{U}^{-1}. \tag{6.28}$$

Similarly, for $\mathbf{A}$ nonsingular

$$\mathbf{A}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^{-1}, \tag{6.29}$$

in which case (6.28) also holds for negative integers p. Because $\mathbf{D}$ is diagonal, (6.28) is easy to calculate: $\mathbf{D}^p$ is just $\mathbf{D}$ with its (diagonal) nonzero elements raised to the pth power.

Example 6.11 (continuation of Example 6.2) *From (6.25) and (6.26), equation (6.28) is*

$$\begin{aligned}\mathbf{A}^p &= \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}^p \\ &= \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \begin{bmatrix} 1^p & 0 & 0 \\ 0 & 3^p & 0 \\ 0 & 0 & (-4)^p \end{bmatrix} \frac{1}{70} \begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}.\end{aligned} \tag{6.30}$$

The reader should verify this for $p = -1,\ 1,$ and 2.

Example 6.12 *Economists and others often have occasions to use systems of linear equations known as linear difference equations. An example of a system of first-order linear difference equations is*

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{d} \tag{6.31}$$

for $t = 1, 2, 3, \ldots$, *and where* $\mathbf{A}$, $\mathbf{x}_0$, *and* $\mathbf{d}$ *are known. The problem is to calculate values of* $\mathbf{x}_t$. *From (6.31) we have*

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}(\mathbf{A}\mathbf{x}_{t-2} + \mathbf{d}) + \mathbf{d} = \mathbf{A}^2\mathbf{x}_{t-2} + (\mathbf{A} + \mathbf{I})\,\mathbf{d} = \cdots \\ &= \mathbf{A}^t\mathbf{x}_0 + (\mathbf{A}^{t-1} + \mathbf{A}^{t-2} + \cdots + \mathbf{A} + \mathbf{I})\mathbf{d}.\end{aligned}$$

Then, providing that $\mathbf{A}^k \to \mathbf{0}$ *as* $k \to \infty$ *and that* $(\mathbf{I} - \mathbf{A})^{-1}$ *exists, use can be made of (4.27) to yield*

$$\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0 + (\mathbf{I} - \mathbf{A}^t)\,(\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{d}. \tag{6.32}$$

The similar canonical form provides a method of calculating $\mathbf{A}^t$; *and if, as has already been assumed,* $\mathbf{A}^t \to \mathbf{0}$ *as* $t \to \infty$, *then we see from (6.32) that the corresponding limit for* $\mathbf{x}_t$ *is* $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}$.

An extension of (6.32) as the solution to (6.31) is to adapt linear difference equations of higher than the first order to being first-order equations. For example, suppose we wish to solve the third-order difference equation

$$y_t = \alpha y_{t-1} + \beta y_{t-2} + \gamma y_{t-3} + \delta, \tag{6.33}$$

for $t = 3, 4, \ldots$, *knowing* α, β, γ, *and* δ *and the initial values* y_0, y_1, *and* y_2. *Define* $z_t = y_{t-1}$ *for* $t = 1, 2, \ldots$ *and* $w_t = y_{t-2} = z_{t-1}$ *for* $t = 2, 3, \ldots$, *and observe that using these and (6.33)*

$$\begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ w_{t-1} \end{bmatrix} + \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix}. \tag{6.34}$$

Hence on defining

$$\mathbf{x}_t = \begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \textit{and} \quad \mathbf{d} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix},$$

(6.34) can be expressed in the form of (6.31) and so (6.32) yields its solution, which is the solution of (6.33).

6.5 SYMMETRIC MATRICES

Symmetric matrices have sufficient notable properties in regard to eigenvalues and eigenvectors as to warrant special attention. Furthermore, the widespread use of symmetric matrices in statistics, through their involvement in quadratic forms (see Section 5.6), make it worthwhile to discuss these properties in detail. Only real, symmetric matrices are considered.

6.5.1 Eigenvalues All Real

The eigenvalues of a matrix of order $n \times n$ are roots of a polynomial equation of degree n and so are not necessarily real numbers. Some may be complex numbers, occurring in pairs as a + ib and its complex conjugate $a - ib$, where $i = \sqrt{-1}$ and a and b are real. However, when **A** is symmetric (and has elements that are real numbers) then all its eigenvalues are real. We state this formally as a lemma.

Lemma 6.1 *The eigenvalues of every real symmetric matrix are real.*

Proof.[1] Suppose λ is a complex eigenvalue of the symmetric matrix $\mathbf{M}_{n\times n}$ with $\mathbf{u}$ being a corresponding eigenvector. If $\lambda = \alpha + i\beta$, define $\bar{\lambda} = \alpha - i\beta$, and for $\mathbf{u} = \{u_k\} = \mathbf{a} + i\mathbf{b}$ define $\bar{\mathbf{u}} = \{\bar{u}_k\} = \mathbf{a} - i\mathbf{b}$. Then by definition, $\mathbf{Mu} = \lambda\mathbf{u}$ so that $\bar{\mathbf{u}}'\mathbf{Mu} = \bar{\mathbf{u}}'\lambda\mathbf{u} = \lambda\bar{\mathbf{u}}'\mathbf{u}$. But $\mathbf{M}\bar{\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$ so that we also have $\bar{\mathbf{u}}'\mathbf{Mu} = (\mathbf{M}\bar{\mathbf{u}})'\,\mathbf{u} = (\bar{\lambda}\bar{\mathbf{u}})'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$. Equating these two expressions for $\bar{\mathbf{u}}'\mathbf{Mu}$ gives $\lambda\bar{\mathbf{u}}'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$ and since $\bar{\mathbf{u}}'\mathbf{u}$ is a sum of squares of real numbers, it is nonzero and so $\lambda = \bar{\lambda}$, that is, $a + ib = a - ib$ and hence $b = 0$. Thus every λ is real. ■

This lemma means that when dealing with symmetric matrices, all eigenvalues are real. Corresponding to each eigenvalue is an eigenvector **u**, say, that is real.

6.5.2 Symmetric Matrices Are Diagonable

For every eigenvalue of a symmetric matrix **A**, condition (6.24) is satisfied: $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for every λ_k (Proof is given in Section 7.1). Therefore, $\mathbf{A} = \mathbf{A}'$ is diagonable. Hence, for any symmetric matrix **A**, we have $\mathbf{U}^{-1}$ of $\mathbf{AU} = \mathbf{UD}$ existing, and $\mathbf{A} = \mathbf{UDU}^{-1}$ and $\mathbf{A}^p = \mathbf{UD}^p\mathbf{U}^{-1}$. Moreover, as is now developed, **U** is orthogonal.

6.5.3 Eigenvectors Are Orthogonal

Symmetric matrices have eigenvectors that are orthogonal to one another. We establish this in two cases: (1) eigenvectors corresponding to different eigenvalues, and (2) eigenvectors corresponding to a multiple eigenvalue.

1. Different Eigenvalues Consider two different eigenvalues $\lambda_1 \neq \lambda_2$ with $\mathbf{u}_1$, $\mathbf{u}_2$ being corresponding eigenvectors. Then with $\mathbf{A} = \mathbf{A}'$ and $\mathbf{Au}_k = \lambda_k\mathbf{u}_k$ we have

$$\lambda_1\mathbf{u}_2'\mathbf{u}_1 = \mathbf{u}_2'\lambda_1\mathbf{u}_1 = \mathbf{u}_2'\mathbf{Au}_1 = \mathbf{u}_1'\mathbf{A}'\mathbf{u}_2 = \mathbf{u}_1'\mathbf{Au}_2 = \mathbf{u}_1'\lambda_2\mathbf{u}_2 = \lambda_2\mathbf{u}_1'\mathbf{u}_2 = \lambda_2\mathbf{u}_2'\mathbf{u}_1;$$

that is, $\lambda_1\mathbf{u}_2'\mathbf{u}_1 = \lambda_2\mathbf{u}_2'\mathbf{u}_1$. But $\lambda_1 \neq \lambda_2$. Therefore $\mathbf{u}_2'\mathbf{u}_1 = 0$; that is, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. Hence, eigenvectors corresponding to different eigenvalues are orthogonal.

[1] Thanks go to G. P. H. Styan for the brevity of this proof.

Before discussing case (2), the following lemma is needed:

Lemma 6.2 *Let $\boldsymbol{B}$ be an $n \times n$ matrix. Solutions to $\boldsymbol{Bx} = \boldsymbol{0}$ can always be found that are orthogonal to one another.*

Proof. Suppose that $\boldsymbol{B}$ has rank r. Then, $\mathbf{Bx} = \mathbf{0}$ has $n - r$ linearly independent solutions. Suppose $\mathbf{x}_1$ is a solution. Consider the equations in $\mathbf{x}$:

$$\begin{bmatrix} \mathbf{B} \\ \mathbf{x}_1' \end{bmatrix} \mathbf{x} = \mathbf{0}. \tag{6.35}$$

Because $\mathbf{x}_1$ is a solution to $\mathbf{Bx} = \mathbf{0}$, $\boldsymbol{x}_1'$ is orthogonal to the rows of $\boldsymbol{B}$ and hence linearly independent of these rows. Therefore, (6.35) has $n - r - 1$ linearly independent solutions. Suppose $\mathbf{x}_2$ is such a solution. Then clearly $\mathbf{x}_2$ satisfies $\mathbf{Bx} = \mathbf{0}$ and is also orthogonal to $\mathbf{x}_1$. Similarly, if $n - r > 2$ a third solution can be obtained by solving

$$\begin{bmatrix} \mathbf{B} \\ \mathbf{x}_1' \\ \mathbf{x}_2' \end{bmatrix} \mathbf{x} = \mathbf{0}$$

which has $n - r - 2$ linearly independent solutions for $\mathbf{x}$. This process can be continued until $n - r$ solutions have been obtained, all orthogonal to one another. ■

2. Multiple Eigenvalues If $\mathbf{A} = \mathbf{A}'$ has λ_k as an eigenvalue with multiplicity m_k, then $\mathbf{A} - \lambda_k\mathbf{I}$ has rank $n - m_k$ and is singular; and there are m_k eigenvectors corresponding to λ_k which are linearly independent solutions to $(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}$ that are also orthogonal to one another by Lemma 6.2. Hence, if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$, there exist m_k linearly independent orthogonal eigenvectors of $\mathbf{A}$ corresponding to λ_k.

Theorem 6.2 (The Spectral Decomposition Theorem) *Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$. There exists an orthogonal matrix $\boldsymbol{P}$ such that*

$$\boldsymbol{A} = \boldsymbol{P\Lambda P}', \tag{6.36}$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{A}$, and the columns of $\boldsymbol{P}$ are eigenvectors of $\mathbf{A}$ that correspond to $\lambda_1, \lambda_2, \dots, \lambda_n$ in the following manner: If $\boldsymbol{P}$ is partitioned as

$$\boldsymbol{P} = [\boldsymbol{p}_1\, \boldsymbol{p}_2\, \cdots\, \boldsymbol{p}_n],$$

where $\boldsymbol{p}_i$ is the ith *column of $\boldsymbol{P}$, then $\boldsymbol{p}_i$ is an eigenvector of $\mathbf{A}$ corresponding to λ_i ($i = 1, 2, \dots, n$). The matrix $\mathbf{A}$ can then be expressed as*

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i'. \tag{6.37}$$

Proof. We have established that $\mathbf{A} = \mathbf{A}'$ is diagonable, that eigenvectors corresponding to different eigenvalues are orthogonal, and that m_k linearly independent eigenvectors corresponding to any eigenvalue λ_k of multiplicity m_k can be obtained such that they are orthogonal. Thus the m_k eigenvectors corresponding to λ_k are orthogonal, not only to one another, but also to the m_l eigenvectors corresponding to each other eigenvalue λ_l; and this is true for every eigenvalue λ_k. Hence eigenvectors for a symmetric matrix can always be found such that they are *all* orthogonal to one another. On normalizing each vector (see Section 5.4) by changing $\boldsymbol{u}$ to $(1/\sqrt{\mathbf{u}'\mathbf{u}})\mathbf{u}$ and arraying the normalized vectors in a matrix $\boldsymbol{P}$, we then have $\boldsymbol{P}$ as an orthogonal matrix and hence

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \boldsymbol{\Lambda} \quad \text{with} \quad \mathbf{P}\mathbf{P}' = \mathbf{I}.$$

It follows that

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'. \tag{6.38}$$

■

The proof of the next theorem can be found in Harville (1997, Theorem 14.5.16, p. 231).

Theorem 6.3 (The Cholesky Decomposition) *Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$.*

(a) *If $\boldsymbol{A}$ is positive definite, then there exists a unique upper triangular matrix $\boldsymbol{T}$ with positive diagonal elements such that*

$$\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}.$$

(b) *If $\boldsymbol{A}$ is non-negative definite with rank equal to r, then there exists a unique upper triangular matrix $\boldsymbol{U}$ with r positive diagonal elements and with $n - r$ zero rows such that*

$$\boldsymbol{A} = \boldsymbol{U}'\boldsymbol{U}.$$

Example 6.13

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

has a characteristic equation which reduces to $(\lambda + 1)^2(\lambda - 5) = 0$. Hence, $\lambda_1 = 5$ with $m_1 = 1$, and $\lambda_2 = -1$ with $m_2 = 2$. For $\lambda_1 = 5$, we get $\boldsymbol{u}_1 = (1, 1, 1)'$ as an eigenvector. For $\lambda_2 = -1$ we find $\boldsymbol{u}_2 = (-2, 1, 1)'$ as an eigenvector that is orthogonal to $\boldsymbol{u}_1$. Since $m_2 = 2$ there should be another eigenvector that is orthogonal to $\boldsymbol{u}_2$ by Case 2 in Section 6.5.3. Such an eigenvector can be obtained by finding a vector $\mathbf{v}_2$ such that $\mathbf{v}_2'\boldsymbol{u}_2 = 0$, that is, $v_{21}(-2) + v_{22}(1) + v_{23}(1) = 0$, where the v_{2i}'s are the elements of $\boldsymbol{v}_2$. We find that $\boldsymbol{u}_3 =$

$(0, -1, 1)'$ *satisfies this condition and is orthogonal to* $\boldsymbol{u}_1$. *Arraying the normalized forms of these vectors as a matrix gives*

$$\boldsymbol{P} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \\ \sqrt{2} & 1 & \sqrt{3} \end{bmatrix}.$$

The reader should verify that $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} = diag\,(5,\ -1,\ -1)$ *and* $\mathbf{PP}' = \mathbf{I}$.

6.5.4 Rank Equals Number of Nonzero Eigenvalues for a Symmetric Matrix

Define $z_{\mathbf{A}}$ as the number of zero eigenvalues of a symmetric matrix $\boldsymbol{A}$ of order $n \times n$; then $n - z_{\mathbf{A}}$ is the number of nonzero eigenvalues. Since, for $\boldsymbol{A}$ being symmetric, $\mathbf{A} = \boldsymbol{P\Lambda P}'$ for nonsingular (orthogonal) $\boldsymbol{P}$, the ranks of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are equal, $r_{\mathbf{A}} = r_{\boldsymbol{\Lambda}}$. But, the only nonzero elements in the diagonal matrix $\boldsymbol{\Lambda}$ are the nonzero eigenvalues, and so its rank is the number of such eigenvalues, $n - z_{\mathbf{A}}$. Hence

$$r_{\mathbf{A}} = n - z_{\mathbf{A}}, \quad \text{for} \quad \mathbf{A} = \mathbf{A}'; \tag{6.39}$$

that is, for symmetric matrices rank equals the number of nonzero eigenvalues.

This result is true not only for all symmetric matrices (because they are diagonable), but also for all diagonable matrices. Other than diagonability there is nothing inherent in the development of (6.39) that uses the symmetry of $\boldsymbol{A}$. Nevertheless, (6.39) is of importance because it applies to all symmetric matrices.

6.6 EIGENVALUES OF ORTHOGONAL AND IDEMPOTENT MATRICES

Theorem 6.4 *If* $\boldsymbol{A}$ *is an* $n \times n$ *orthogonal matrix and* λ *is an eigenvalue of* $\boldsymbol{A}$, *then* $\lambda = \pm 1$.

Proof. We have that $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, where $\boldsymbol{x}$ is an eigenvector corresponding to λ. Since $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$, multiplying both sides on the left by $\boldsymbol{x}'$ and on the right by $\boldsymbol{x}$ we get $\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{Ax} = \boldsymbol{x}'\boldsymbol{x}$. This implies that $\lambda^2\boldsymbol{x}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x}$. We conclude that $\lambda^2 = 1$, that is, $\lambda = \pm 1$ since $\boldsymbol{x} \neq \mathbf{0}$. ■

Theorem 6.5 *If* $\boldsymbol{B}$ *is an* $n \times n$ *idempotent matrix, then its eigenvalues are equal to 0 or 1. Furthermore, if the rank of* $\boldsymbol{B}$ *is r, then* $\boldsymbol{B}$ *has r eigenvalues all equal to 1.*

Proof. Let λ be an eigenvalue of $\boldsymbol{B}$ with an eigenvector $\boldsymbol{x}$. Then, $\boldsymbol{B}^2\boldsymbol{x} = \boldsymbol{Bx} = \lambda\boldsymbol{x}$. But, $\boldsymbol{B}^2\boldsymbol{x} = \boldsymbol{BBx} = \boldsymbol{B}\lambda\boldsymbol{x} = \lambda\boldsymbol{Bx} = \lambda^2\boldsymbol{x}$. We conclude that $\lambda\boldsymbol{x} = \lambda^2\boldsymbol{x}$. Hence, $\lambda = \lambda^2$, that is, $\lambda = 0$ or 1 since $\boldsymbol{x} \neq \mathbf{0}$. Furthermore, since $\boldsymbol{B}$ is idempotent, then by Theorem 5.1, $r(\boldsymbol{B}) = tr(\boldsymbol{B})$. But, $r = r(\boldsymbol{B})$ and $tr(\boldsymbol{B})$ is the sum of the nonzero eigenvalues, which are equal to 1, as the remaining ones are equal to zero, we conclude that $\boldsymbol{B}$ has r nonzero eigenvalues that are equal to 1. ■

6.6.1 Eigenvalues of Symmetric Positive Definite and Positive Semidefinite Matrices

Theorem 6.6 *The symmetric matrix* $\mathbf{A}$ *is positive definite if and only if its eigenvalues are positive.*

Proof. Let λ and $\boldsymbol{x}$ be an eigenvalue and corresponding eigenvector of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}'\boldsymbol{x}$. Since $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, then $\lambda > 0$. Vice versa, if the eigenvalues of $\boldsymbol{A}$ are positive, then by the spectral decomposition theorem (Theorem 6.2), $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{x}$, where $\boldsymbol{P}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$. Since the diagonal elements of $\boldsymbol{\Lambda}$ are all positive, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x}$. ■

The proof of the following theorem is left as an exercise.

Theorem 6.7 *Let* $\mathbf{A}$ *be a symmetric matrix. Then,* $\mathbf{A}$ *is positive semidefinite if and only if its eigenvalues are non-negative with at least one equal to zero.*

The proof of the next theorem can be found in Banerjee and Roy (2014, Theorem 13.18).

Theorem 6.8 *Let* $\mathbf{A}$ *be a symmetric matrix. Then,* $\mathbf{A}$ *is non-negative definite if and only if all its principal minors (see Section 3.6) are non-negative.*

Example 6.14 *The characteristic equation for the positive semidefinite matrix of Section 5.7*

$$\mathbf{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \quad \text{is} \quad \lambda(\lambda - 14)(\lambda - 53) = 0, \tag{6.40}$$

so that the eigenvalues are 0, 14, and 53.

A further feature of non-negative definite matrices (see Section 5.7) is that each is the square of some other matrix of real elements. For if $\mathbf{A}$ *is non-negative definite we have, by the Spectral Decomposition Theorem,* $\mathbf{A} = \boldsymbol{PDP}'$ *where*

$$\boldsymbol{D} = diag\left\{\lambda_l\right\} = \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} \quad for \quad \mathbf{D}^{\frac{1}{2}} = diag\{\sqrt{\lambda_l}\}$$

with $\sqrt{\lambda_l}$ *being the positive square root. Hence,*

$$\mathbf{A} = \mathbf{PDP}' = \mathbf{PD}^{\frac{1}{2}}\mathbf{P}'\mathbf{PD}^{\frac{1}{2}}\mathbf{P}' = \mathbf{H}^2 \quad for \quad \mathbf{H} = \mathbf{PD}^{\frac{1}{2}}\mathbf{P}'. \tag{6.41}$$

For positive definite matrices no λ_l *is zero and so* $\mathbf{D}, \mathbf{D}^{\frac{1}{2}}$, *and* $\mathbf{H}$ *are all of full rank giving* $\mathbf{A}^{-1} = \left(\mathbf{H}^{-1}\right)^2$.

Also, if $\mathbf{P}'\mathbf{AP} = \mathbf{D} = \begin{bmatrix} \boldsymbol{\Delta}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ *where* $\Delta = diag\{\sqrt{\lambda_i}\}$ *for just the* $r_\mathbf{A}$ *nonzero eigenvalues, then with*

$$\mathbf{S} = \begin{bmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r_A} \end{bmatrix} \text{ and } \mathbf{T} = \mathbf{PS}, \text{ we have } \mathbf{T}'\mathbf{AT} = \begin{bmatrix} \mathbf{I}_{r_A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{6.42}$$

where $\mathbf{T}$ *is nonsingular.*

Example 6.15 (continued from Example 6.14) *Eigenvectors of* $\mathbf{A}$ *in (6.40) corresponding to the eigenvalues 14, 53, and 0 are*

$$\begin{bmatrix} 2 \\ -13 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \tag{6.43}$$

respectively. Hence $\mathbf{H}$ *of (6.41) is*

$$\mathbf{H} = \begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix}$$
$$\times \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{53} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix} \begin{bmatrix} 2 & -13 & 3 \\ -3 & 0 & 2 \\ 2 & 1 & 3 \end{bmatrix},$$

with the second and fourth matrices in this product representing the normalization of the eigenvectors in (6.43). After simplification, $\mathbf{H}$ *can be written as*

$$\mathbf{H} = \frac{1}{13\sqrt{14}} \begin{bmatrix} 4+9a & -26 & 6-6a \\ -26 & 169 & -39 \\ 6-6a & -39 & 9+4a \end{bmatrix} \quad \text{with} \quad a = \sqrt{14(53)}.$$

The reader can verify that $\mathbf{H}^2 = \mathbf{A}$, *as in (6.41). With*

$$\mathbf{S} = \begin{bmatrix} 1/\sqrt{14} & 0 & 0 \\ 0 & 1/\sqrt{53} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{T} = \mathbf{PS} = \begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{14(182)} & 0 & 0 \\ 0 & 1/\sqrt{53(13)} & 0 \\ 0 & 0 & \sqrt{1/14} \end{bmatrix},$$

(6.42) can also be verified.

Theorem 6.9 *Let $\mathbf{A}$ be an $n \times n$ symmetric matrix of rank r. Then $\mathbf{A}$ can be written as*

$$\mathbf{A} = \mathbf{L}\mathbf{L}', \tag{6.44}$$

where $\mathbf{L}$ is $n \times r$ of rank r, that is, of $\mathbf{L}$ is of full column rank.

Proof. By the spectral decomposition theorem (Theorem 6.2), $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix of corresponding eigenvectors of $\boldsymbol{A}$. The matrix $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{D}_r, \mathbf{0}), \tag{6.45}$$

where $\boldsymbol{D}_r$ is a diagonal matrix of order $r \times r$ whose diagonal elements are the nonzero eigenvalues of $\boldsymbol{A}$, and the first r columns of $\boldsymbol{P}$ are the corresponding eigenvectors. we then have

$$\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D}_r^{1/2} \\ \mathbf{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{D}_r^{1/2} & \mathbf{0}' \end{bmatrix}\boldsymbol{P}' = \boldsymbol{L}\boldsymbol{L}', \tag{6.46}$$

where

$$\boldsymbol{L} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D}_r^{1/2} \\ \mathbf{0} \end{bmatrix}.$$

Note that $\boldsymbol{L}$ is of order $n \times r$ and rank r. Its elements are not necessarily real since some of the the diagonal elements of $\boldsymbol{D}_r$ may be negative. ■

Corollary 6.1 *If $\mathbf{A}$ is a symmetric positive semidefinite matrix of order $n \times n$ and rank r, then it can be written as $\mathbf{A} = \mathbf{K}\mathbf{K}'$ where $\mathbf{K}$ is a real matrix of order $n \times r$ and rank r.*

Proof. This result follows from Theorem 6.7 since $\boldsymbol{A}$, being positive semidefinite, its nonzero eigenvalues are positive. Hence, the diagonal elements of the diagonal matrix $\boldsymbol{D}_r$ are all positive which makes the matrix $\boldsymbol{L}$ a real matrix. In this case, $\boldsymbol{K} = \boldsymbol{L}$. ■

Corollary 6.2 *If $\mathbf{A}$ is a symmetric positive definite matrix of order $n \times n$, then $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{M}\mathbf{M}'$, where $\mathbf{M}$ is a nonsingular matrix.*

Proof. This follows directly from Theorem 6.6 since the diagonal elements of the diagonal matrix $\boldsymbol{\Lambda}$ are positive. Hence, $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{M}'$, where $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}$, which is a nonsingular matrix. ■

6.7 EIGENVALUES OF DIRECT PRODUCTS AND DIRECT SUMS OF MATRICES

Theorem 6.10 *Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of orders $m \times m$ and $n \times n$, respectively. Let λ and $\mathbf{x}$ be an eigenvalue and a corresponding eigenvector of $\mathbf{A}$. Likewise, let μ and $\mathbf{y}$ denote* *an eigenvalue and a corresponding eigenvector of* $\boldsymbol{B}$. *Then,* $\lambda\mu$ *is an eigenvalue of* $\boldsymbol{A} \otimes \boldsymbol{B}$ *with a corresponding eigenvector* $\boldsymbol{x} \otimes \boldsymbol{y}$.

Proof. We have that $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ and $\boldsymbol{By} = \mu\boldsymbol{y}$. Hence,

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{x} \otimes \boldsymbol{y}) = (\boldsymbol{Ax}) \otimes (\boldsymbol{By}) = (\lambda\boldsymbol{x}) \otimes (\mu\boldsymbol{y}) = \lambda\mu(\boldsymbol{x} \otimes \boldsymbol{y}).$$

This shows that $\lambda\mu$ is an eigenvalue of $\boldsymbol{A} \otimes \boldsymbol{B}$ with an eigenvector $\boldsymbol{x} \otimes \boldsymbol{y}$. ∎

Based on this theorem it can be seen that if $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvalues of $\boldsymbol{A}$, and $\mu_1, \mu_2, \ldots, \mu_n$ are eigenvalues of $\boldsymbol{B}$, then $\lambda_i\mu_j$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$ are eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$.

Corollary 6.3 *The rank of* $\boldsymbol{A} \otimes \boldsymbol{B}$ *is the product of the ranks of* $\boldsymbol{A}$ *and* $\boldsymbol{B}$.

Proof. From Section 4.15 (item xii), we can write $r(\boldsymbol{A} \otimes \boldsymbol{B}) = r\{(\boldsymbol{A} \otimes \boldsymbol{B})'(\boldsymbol{A} \otimes \boldsymbol{B})\} = r\{(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})\}$. Clearly, $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$ is symmetric. Thus, by Section 6.5.4, its rank is equal to the number of nonzero eigenvalues of $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$, which are real scalars. If τ_i is the ith eigenvalue of $\boldsymbol{A}'\boldsymbol{A}$ and κ_j is the jth eigenvalue of $\boldsymbol{B}'\boldsymbol{B}$, then by Theorem 6.7 $\tau_i\kappa_j$ is an eigenvalue of $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$, which is nonzero if and only if both τ_i and κ_j are nonzero. It follows that the number of nonzero eigenvalues of $(\boldsymbol{A}'\boldsymbol{A}) \times (\boldsymbol{B}'\boldsymbol{B})$ is the product of the nonzero eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$, which is equal to the rank of $\boldsymbol{A}'\boldsymbol{A}$, or equivalently, the rank of $\boldsymbol{A}$ by Section 4.15 (item xii), and the number of nonzero eigenvalues of $\boldsymbol{B}'\boldsymbol{B}$, which is the rank of $\boldsymbol{B}'\boldsymbol{B}$, or the rank of $\boldsymbol{B}$. We then conclude that

$$r(\boldsymbol{A} \otimes \boldsymbol{B}) = r(\boldsymbol{A})r(\boldsymbol{B}).$$

∎

Corollary 6.3 confirms the result given in Section 4.15 (item xv).

Theorem 6.11 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be matrices of orders* $m \times m$ *and* $n \times n$*, respectively. Let* λ *be an eigenvalue of* $\boldsymbol{A}$ *with a corresponding eigenvector* $\boldsymbol{x}$*, and let* μ *be an eigenvalues of* $\boldsymbol{B}$ *with a corresponding eigenvector* $\boldsymbol{y}$. *Then,* $\lambda \oplus \mu$ *is an eigenvalue of* $\boldsymbol{A} \oplus \boldsymbol{B}$ *with a corresponding eigenvector* $\boldsymbol{x} \oplus \boldsymbol{y}$.

Proof.

$$\begin{aligned}(\boldsymbol{A} \oplus \boldsymbol{B})(\boldsymbol{x} \oplus \boldsymbol{y}) &= (\boldsymbol{Ax}) \oplus (\boldsymbol{By}) \\ &= (\lambda\boldsymbol{x}) \oplus (\mu\boldsymbol{y}) \\ &= (\boldsymbol{x} \oplus \boldsymbol{y})(\lambda \oplus \mu).\end{aligned}$$

We conclude that $\lambda \oplus \mu$ is eigenvalue of $\boldsymbol{A} \oplus \boldsymbol{B}$ with a corresponding eigenvector $\boldsymbol{x} \oplus \boldsymbol{y}$. Thus, if $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvectors of $\boldsymbol{A}$, and $\mu_1, \mu_2, \ldots, \mu_n$ are eigenvectors of $\boldsymbol{B}$, then $\lambda_1, \lambda_2, \ldots, \lambda_m; \mu_1, \mu_2, \ldots, \mu_n$ are eigenvalues of $\boldsymbol{A} \oplus \boldsymbol{B}$. ∎

Based on this theorem it is easy to see that $tr(\boldsymbol{A} \oplus \boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$.

6.8 NONZERO EIGENVALUES OF *AB* AND *BA*

The following theorem relates the nonzero eigenvalues of $\boldsymbol{AB}$ to those of $\boldsymbol{BA}$. It has useful applications in statistics as will be seen later:

Theorem 6.12 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be matrices of orders* $m \times n$ *and* $n \times m$*, respectively* ($n \geq m$)*. Let* $\lambda \neq 0$*. Then,*

(a) $|\boldsymbol{BA} - \lambda \boldsymbol{I}_n| = (-\lambda)^{n-m} |\boldsymbol{AB} - \lambda \boldsymbol{I}_m|$.

(b) *The nonzero eigenvalues of* $\mathbf{BA}$ *are the same as those of* $\mathbf{AB}$*. Furthermore, the multiplicity of a nonzero eigenvalue of* $\mathbf{BA}$ *is the same as when regarded as an eigenvalue of* $\mathbf{AB}$*.*

(c) *If* $m = n$*, then all the eigenvalues of* $\mathbf{BA}$ *(not just the nonzero ones) are the same as those of* $\mathbf{AB}$*.*

Proof.

(a) Consider the determinant of the matrix $\boldsymbol{L}$ partitioned as

$$|\boldsymbol{L}| = \begin{vmatrix} \boldsymbol{I}_m & (1/\lambda)\boldsymbol{A} \\ \boldsymbol{B} & \boldsymbol{I}_n \end{vmatrix}, \quad \lambda \neq 0. \tag{6.47}$$

By applying property (g) in Section 4.12, the determinant of $\boldsymbol{L}$ can be written as

$$\begin{aligned} |\boldsymbol{L}| &= |\boldsymbol{I}_n - \boldsymbol{B}(1/\lambda)\boldsymbol{A}|, \text{ since here } \boldsymbol{A}_{11} = \boldsymbol{I}_m, \;\; \lambda \neq 0, \\ |\boldsymbol{L}| &= |\boldsymbol{I}_m - (1/\lambda)\boldsymbol{AB}|, \text{ since here } \boldsymbol{A}_{22} = \boldsymbol{I}_n, \;\; \lambda \neq 0. \end{aligned}$$

We conclude that

$$|\boldsymbol{I}_n - \boldsymbol{B}(1/\lambda)\boldsymbol{A}| = |\boldsymbol{I}_m - (1/\lambda)\boldsymbol{AB}|, \;\; \lambda \neq 0.$$

Hence,

$$\lambda^{-n} |\lambda \boldsymbol{I}_n - \boldsymbol{BA}| = \lambda^{-m} |\lambda \boldsymbol{I}_m - \boldsymbol{AB}|, \;\; \lambda \neq 0.$$

This can be written as

$$|\boldsymbol{BA} - \lambda \boldsymbol{I}_n| = (-\lambda)^{n-m} |\boldsymbol{AB} - \lambda \boldsymbol{I}_m|, \;\; \lambda \neq 0. \tag{6.48}$$

(b) It can be seen from (6.48) that the nonzero eigenvalues of $\boldsymbol{BA}$ are the same as those of $\boldsymbol{AB}$. Furthermore, each nonzero eigenvalue of $\boldsymbol{BA}$ has the same multiplicity as when regarded as a nonzero eigenvalue of $\boldsymbol{AB}$.

(c) This is obvious from part (a). ∎

REFERENCES

Banerjee, S. and Roy, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, Boca Raton, FL.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

EXERCISES

6.1 Find the eigenvalues and eigenvectors of the following symmetric matrices. In each case combine the eigenvectors into an orthogonal matrix $\boldsymbol{P}$ and verify that $\mathbf{P}'\mathbf{A}\mathbf{P} = \boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues.

$$\mathbf{B} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix} \mathbf{C} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & -4 \\ 1 & -4 & 1 \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix} \mathbf{F} = \begin{bmatrix} -9 & 2 & 6 \\ 2 & -9 & 6 \\ 6 & 6 & 7 \end{bmatrix}$$

6.2 Show that if $\mathbf{B} = \mathbf{A}^2 + \mathbf{A}$, and if λ is an eigenvalue of $\mathbf{A}$ then $\lambda^2 + \lambda$ is an eigenvalue of $\mathbf{B}$.

6.3 Show that the eigenvectors of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{are of the form} \quad \begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -b \\ a - \lambda_2 \end{bmatrix}$$

where λ_1 and λ_2 are the eigenvalues. Verify that $2a = \lambda_1 + \lambda_2$ when $a = d$.

6.4 Factorize the characteristic equation of

$$\mathbf{A} = \begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix}$$

using the eigenvalue $\lambda = a - b$.

6.5 **(a)** Find the eigenvalues of $\mathbf{I} - \boldsymbol{xx}'$ for $\boldsymbol{x}$ being a vector of order $n \times 1$.

(b) Find orthogonal eigenvectors of $\mathbf{I} - \boldsymbol{xx}'$.

(c) Let $\boldsymbol{\Lambda}$ be the diagonal matrix of eigenvalues of $\boldsymbol{I} - \boldsymbol{xx}'$ and let $\boldsymbol{P}$ be the corresponding orthogonal matrix of eigenvectors. Find these matrices and show that $\boldsymbol{I} - \boldsymbol{xx}' = \boldsymbol{P\Lambda P}'$.

6.6 Using the characteristic equation, show for λ being an eigenvalue of $\boldsymbol{A}$ that $1/(1 + \lambda)$ is an eigenvalue of $(\mathbf{I} + \boldsymbol{A})^{-1}$.

6.7 When eigenvalues of $\mathbf{A}$ are positive, prove that those of $\mathbf{A} + \mathbf{A}^{-1}$ are equal or greater than 2.

6.8 For $\mathbf{A}$ being symmetric of order $n \times n$, with eigenvalues λ_ι, prove that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = \sum_{k=1}^{n} \lambda_k^2.$$

6.9 Show that if λ is an eigenvalue of an orthogonal matrix, then so is $1/\lambda$.

6.10 Prove that the eigenvalues of $\boldsymbol{BA}$ are the same as those of $\boldsymbol{ABA}$ when $\boldsymbol{A}$ is idempotent.

6.11 Suppose $\boldsymbol{T}$ and $\boldsymbol{K}$ commute in multiplication. For $\boldsymbol{x}$ being an eigenvector of $\boldsymbol{T}$, show that $\boldsymbol{Kx}$ is also.

6.12 The eigenvalues of $\mathbf{A}$ with respect to $\mathbf{V}$ are defined as the solutions for λ to $|\mathbf{A} - \lambda\mathbf{V}| = 0$; and then $\mathbf{At} = \lambda\mathbf{Vt}$ defines $\mathbf{t}$ as an eigenvector of $\mathbf{A}$ with respect to $\mathbf{V}$. If $\mathbf{V}$ is positive definite, symmetric,

(a) Show that λ is an eigenvalue of $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$.

(b) Find $\mathbf{t}$ in terms of an eigenvector of $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$.

6.13 For any idempotent matrix $\mathbf{A}$, prove that

(a) $\mathbf{A}^k$ has the same eigenvalues as $\mathbf{A}$,

(b) $\mathbf{A}^k$ has rank $r_{\mathbf{A}}$

and

(c) $r(\mathbf{I} - \mathbf{A}) = n - r_{\mathbf{A}}$ for $\mathbf{A}$ of order $n \times n$.

6.14 Let $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{Y}$ be matrices of order $n \times n$ such that $\boldsymbol{Q} = \boldsymbol{RR}'$. Prove that the eigenvalues of $\boldsymbol{YQ}$ are also eigenvalues of $\boldsymbol{R}'\boldsymbol{YR}$.

6.15 Find the eigenvalues of $a\mathbf{I} + b\mathbf{J}$ where $\boldsymbol{J}$ has order $r \times r$ and every element is equal to one. Find also the corresponding eigenvectors.

6.16 Let $\boldsymbol{A}$ be an $n \times n$ matrix and $\boldsymbol{C}$ be a nonsingular matrix of the same order. Show that $\boldsymbol{A}$, $\boldsymbol{C}^{-1}\boldsymbol{AC}$, and $\boldsymbol{CAC}^{-1}$ have the same set of eigenvalues.

6.17 It might seem "obvious" that all eigenvalues of a matrix being zero implies the matrix is null.

(a) Prove this, for real, symmetric matrices.

(b) Create a numerical example (e.g., a 2×2 matrix) illustrating that the statement is not true for nonsymmetric matrices. Note that your example also illustrates that Section 6.5.4 is not valid for nonsymmetric matrices.

6.18 Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix.

(a) Show that

$$\lambda_{min} \leq \frac{\boldsymbol{x}'\boldsymbol{Ax}}{\boldsymbol{x}'\boldsymbol{x}} \leq \lambda_{max}, \quad \text{for any } \boldsymbol{x} \neq \mathbf{0},$$

where λ_{min} is the smallest eigenvalue of $\boldsymbol{A}$ and λ_{max} is the largest eigenvalue.

(b) For what values of $\boldsymbol{x}$ can the lower and upper bound in part (a) be achieved?

Note: The ratio $\boldsymbol{x}'\boldsymbol{Ax}/\boldsymbol{x}'\boldsymbol{x}$ is called *Rayleigh's quotient*.

6.19 If $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$, λ_{min} and λ_{max} are the same as in Exercise 6.18, then

$$\lambda_{min} \leq \frac{1}{n} \sum_{ij} a_{ij} \leq \lambda_{max},$$

where a_{ij} is the (ij)th element of $\boldsymbol{A}$.

6.20 Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$. Show that $\boldsymbol{A}$ can be written in the form:

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{A}_i,$$

where $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ are idempotent matrices of rank 1, $\boldsymbol{A}_i \boldsymbol{A}_j = \boldsymbol{0}$, $i \neq j$, and λ_i is the ith eigenvalue of $\boldsymbol{A}$, $i = 1, 2, \ldots, n$.

6.21 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ symmetric matrices with $\boldsymbol{A}$ non-negative definite.Show that

$$e_{min}(\boldsymbol{B}) tr(\boldsymbol{A}) \leq tr(\boldsymbol{AB}) \leq e_{max}(\boldsymbol{B}) tr(\boldsymbol{A}),$$

where e_{min} and e_{max} are the smallest and largest eigenvalues of $\boldsymbol{A}$, respectively.

6.22 Let $\boldsymbol{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$. Show that

$$e_{min} \leq a_{ii} \leq e_{max}, \quad i = 1, 2, \ldots, n.$$

6.23 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric of the same order and $\boldsymbol{B}$ is positive definite. Prove that solutions for θ to $|\boldsymbol{A} - \theta \boldsymbol{B}| = 0$ are real.

6.24 Let $\boldsymbol{A}$ be positive semidefinite of order $n \times n$. Show that

$$tr(\boldsymbol{A}^2) \leq e_{max}(\boldsymbol{A}) tr(\boldsymbol{A}).$$

6.25 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of orders $n \times n$ and $m \times m$, respectively. The *Kronecker Sum* of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted by $\boldsymbol{A} \uplus \boldsymbol{B}$, is defined as

$$\boldsymbol{A} \uplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I}_m + \boldsymbol{I}_n \otimes \boldsymbol{B}.$$

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be eigenvectors corresponding to the eigenvalues λ and μ of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Show that $\lambda + \mu$ is an eigenvalue of $\boldsymbol{A} \uplus \boldsymbol{B}$ with an eigenvector $\boldsymbol{x} \otimes \boldsymbol{y}$.

6.26 Find the eigenvalues and corresponding eigenvectors of $\boldsymbol{A} \uplus \boldsymbol{B}$, where

$$\boldsymbol{A} = \begin{bmatrix} 4 & -2 \\ 5 & -3 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix}.$$

6.27 If the full-rank factorization of $\boldsymbol{A}$ having order $n \times n$ and rank r is $\boldsymbol{BC}$, show that the characteristic equation of $\boldsymbol{A}$ is $(-\lambda)^{n-r} |\boldsymbol{CB} - \lambda \boldsymbol{I}_r|$.

6.28 Let $\boldsymbol{A}$ be matrix of order $n \times n$. Show that $\boldsymbol{A}$ and its transpose, $\boldsymbol{A}'$, have the same set of eigenvalues.

6.29 Prove that the eigenvalues of a skew-symmetric matrix (defined in Section 5.1.6) are zero or imaginary.

7

Diagonalization of Matrices

Eigenvalues and eigenvectors are a foundation for extending both applications and theory far beyond the horizons of this book. A few indications of this are given here. First, with two prerequisite lemmas, comes proof of the diagonability theorem, together with proof that all symmetric matrices are regular, that is, diagonable. Second are some results for the simultaneous diagonalization of symmetric matrices; third is the Cayley–Hamilton theorem and finally, the very useful and important singular value decomposition of a matrix.

Notation: Most of the following notation comes from Chapter 6, but is summarized here for convenience.

$\mathbf{A}$ is square, of order $n \times n$ and rank $r_{\mathbf{A}} = r$.
s = number of distinct eigenvalues.
m_k = multiplicity of the eigenvalue λ_k, with $\sum_{k=1}^{s} m_k = n$.
$z_{\mathbf{A}}$ is the number of zero eigenvalues of $\mathbf{A}$.
$\boldsymbol{D}$ = diagonal matrix of all n eigenvalues.
$\boldsymbol{U}$ = a matrix of eigenvectors used in $\boldsymbol{AU} = \boldsymbol{UD}$.
$\boldsymbol{B} = \boldsymbol{A} - \lambda\boldsymbol{I}$ for λ_k an eigenvalue of $\boldsymbol{A}$.

7.1 PROVING THE DIAGONABILITY THEOREM

7.1.1 The Number of Nonzero Eigenvalues Never Exceeds Rank

It is shown in Section 6.5.4, formula (6.39), that for symmetric matrices rank equals the number of nonzero eigenvalues, $r_{\mathbf{A}} = n - z_{\mathbf{A}}$; and we remarked there that this is so for any

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

diagonable matrix **A**. This is part of a more general result given in the following lemma applicable to any square matrix, whether diagonable or not.

Lemma 7.1 *For* $\mathbf{A}_{n\times n}$ *of rank* $r_{\mathbf{A}}$*, with* $z_{\mathbf{A}}$ *zero eigenvalues,*

$$r_{\mathbf{A}} \geq n - z_{\mathbf{A}}; \tag{7.1}$$

that is, the number of nonzero eigenvalues never exceeds rank.

Proof. Let $r_{\mathbf{A}} = r$. Then, when the full-rank factorization of **A** is **XY**, its characteristic equation can be expressed (Exercise 27 in Chapter 6) as

$$(-\lambda)^{n-r}\, |\mathbf{YX} - \lambda\mathbf{I}| = 0. \tag{7.2}$$

Therefore $\lambda = 0$ is certainly a root $n - r$ times. But to the extent that **YX** has zero eigenvalues, $\lambda = 0$ can be an eigenvalue of **A** more than $n - r$ times. Therefore $z_{\mathbf{A}} \geq n - r_{\mathbf{A}}$. ■

Example 7.1

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{bmatrix} \tag{7.3}$$

has rank 2 (row 3 = twice row 1 + row 2); and its characteristic equation reduces to $\lambda^2(\lambda - 5) = 0$*. Hence* $r_{\mathbf{A}} = 2 > n - z_{\mathbf{A}} = 3 - 2 = 1$*, so illustrating the inequality in (7.1).*

7.1.2 A Lower Bound on $r(\mathbf{A} - \lambda_k\mathbf{I})$

A matrix **A** is diagonable only if $r\left(\mathbf{A} - \lambda_k\mathbf{I}\right) = n - m_k$ for every eigenvalue λ_k of multiplicity m_k. Example 6.10 at the end of Section 6.4 illustrates this condition not being satisfied. For cases like this, the following lemma shows that $r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - m_k$, that is, when $r(\mathbf{A} - \lambda_k\mathbf{I})$ does not equal $n - m_k$, it always exceeds $n - m_k$.

Lemma 7.2 *When* λ_k *is an eigenvalue of* $\mathbf{A}_{n\times n}$ *with multiplicity* m_k*, then*

$$r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - m_k. \tag{7.4}$$

Proof. Define $\mathbf{B} = \mathbf{A} - \lambda_k\mathbf{I}$ and $p_k = n - r_{\mathbf{B}}$. From Lemma 7.1

$$z_{\mathbf{B}} \geq p_k. \tag{7.5}$$

For θ being an eigenvalue of **B**, the corresponding eigenvalue of $\mathbf{A} = \mathbf{B} + \lambda_k\mathbf{I}$ is $\theta + \lambda_k$. Since by (7.5) we have $\theta = 0$ not fewer than p_k times, $0 + \lambda_k = \lambda_k$ is an eigenvalue of **A** not less than p_k times; that is, $m_k \geq p_k$. Hence, $r_{\mathbf{B}} = r(\mathbf{A} - \lambda_k\mathbf{I}) = n - p_k \geq n - m_k$, and (7.4) is established. ■

Corollary 7.1 *For λ_k being a simple root, $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$.*

Proof. Lemma 7.2 gives $r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - 1$. But $|\mathbf{A} - \lambda_k\mathbf{I}| = 0$ and so $r(\mathbf{A} - \lambda_k\mathbf{I}) < n$. Therefore, $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1 = n - m_k$. ∎

7.1.3 Proof of the Diagonability Theorem

For convenience, the theorem is restated.

Theorem 7.1 $\mathbf{A}_{n\times n}$, *having eigenvalues λ_k with multiplicity m_k for $k = 1, 2, \ldots, s$ and $\sum_{k=1}^{s} m_k = n$, has n eigenvectors that are linearly independent if and only if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for all $k = 1, 2, \ldots, s$; whereupon* $\mathbf{U}$ *of* $\mathbf{AU} = \mathbf{UD}$ *is nonsingular and* $\mathbf{A}$ *is diagonable as* $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$.

Proof. (We are indebted to Dr. B. L. Raktoe for the bulk of this proof.) Sufficiency: if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$, then $\mathbf{U}$ is nonsingular.

Since $r(A - \lambda_k\mathbf{I}) = n - m_k$, the equation $(\mathbf{A} - \lambda_k\mathbf{I})\boldsymbol{x} = \mathbf{0}$ has exactly $n - (n - m_k) = m_k$ linearly independent non-null solutions. By definition, these solutions are eigenvectors of $\boldsymbol{A}$. Hence, associated with each λ_k there is a set of m_k linearly independent eigenvectors.

To show that the sets are linearly independent of each other, suppose they are not, and that one vector of the second set of vectors, $\mathbf{y}_2$ say, is a linear combination of vectors of the first set, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{m_1}$. Then $\mathbf{y}_2 = \sum_{i=1}^{m_1} c_i\mathbf{z}_i$ for some scalars c_l not all zero. Multiplying this equation by $\mathbf{A}$ leads to $\mathbf{Ay}_2 = \sum c_l\mathbf{Az}_l$. Because $\mathbf{y}_2$ and the $\mathbf{z}$'s are eigenvectors corresponding respectively to the different eigenvalues λ_2 and λ_1, this means $\lambda_2\mathbf{y}_2 = \sum c_l\lambda_1\mathbf{z}_l = \lambda_1 \sum c_l\mathbf{z}_l = \lambda_1\mathbf{y}_2$, which cannot be true because $\lambda_2 \neq \lambda_1$ and they are not both zero. Therefore, the supposition is wrong, and we conclude that all s sets of m_k eigenvectors, for $k = 1, 2, \ldots, s$ are linearly independent; that is, $\boldsymbol{U}$ is nonsingular.

Necessity: that if $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ exists, then $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$.

Because $\mathbf{D}$ is a diagonal matrix of all n eigenvalues, $(\mathbf{D} - \lambda_k\mathbf{I})$ has exactly m_k zeros in its diagonal and hence $r(\mathbf{D} - \lambda_k\mathbf{I}) = n - m_k$. But $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ so that $\mathbf{A} = \mathbf{UDU}^{-1}$. Therefore, $(\mathbf{A} - \lambda_k\mathbf{I}) = (\mathbf{UDU}^{-1} - \lambda_k\mathbf{I}) = \mathbf{U}(\mathbf{D} - \lambda_k\mathbf{I})\mathbf{U}^{-1}$, and since multiplication by nonsingular matrices does not affect rank, $r(\mathbf{A} - \lambda_k\mathbf{I}) = r(\mathbf{D} - \lambda_k\mathbf{I}) = n - m_k$. ∎

7.1.4 All Symmetric Matrices Are Diagonable

A symmetric matrix is diagonable because, as shown by the following lemma, the condition $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ of the diagonability theorem is satisfied for every eigenvalue of $\mathbf{A} = \mathbf{A}'$.

Lemma 7.3 *If λ_k is an eigenvalue of* $\mathbf{A} = \mathbf{A}'$, *having multiplicity m_k, then $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$.*

Proof. Use $\mathbf{B} = \mathbf{A} - \lambda_k\mathbf{I}$. With $\mathbf{A}$ being symmetric, λ_k is real, and so $\mathbf{B}$ is real; and $\mathbf{B}$ is symmetric. But $r_{\mathbf{B}} = r_{\mathbf{B}'\mathbf{B}}$, and so $r_{\mathbf{B}} = r_{\mathbf{B}^2} = n - p = t$, say, so defining p and t. Therefore, [see Section 4.15, item (iv)] at least one submatrix of order $t \times t$ in $\mathbf{B}$ is nonsingular, and so the corresponding principal minor in $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$ is nonzero. And because (by Exercise 23 in Chapter 5) $\mathbf{B}'\mathbf{B}$ is positive semidefinite, that principal minor of $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$ is positive. Therefore the sum of such principal minors, $\text{tr}_t(\mathbf{B}^2) \neq 0$. Also, because $r_{\mathbf{B}^2} = t$, all minors of order greater than t are zero, and so $\text{tr}_i(\mathbf{B}^2) = 0$ for $i > t$. Hence, the characteristic equation of $\mathbf{B}^2$ is

$$(-\lambda)^n + (-\lambda)^{n-1}\,\text{tr}_1(\mathbf{B}^2) + \cdots + (-\lambda)^{n-1}\,\text{tr}_t\left(\mathbf{B}^2\right) = 0.$$

Because the last term of this equation is nonzero, λ^{n-t} factors out; that is, $\lambda = 0$ is a root $n - t$ times. Thus $z_{\mathbf{B}^2} = n - t = p$ and so $z_{\mathbf{B}} = p$. But for θ being an eigenvalue of $\mathbf{B}$, that of $\mathbf{A} = \mathbf{B} + \lambda_k\mathbf{I}$ is $\theta + \lambda_k$, and so $0 + \lambda_k$ is an eigenvalue of $\mathbf{A}$ with multiplicity p. Hence $p = m$. Thus $r_{\mathbf{B}} = r(\mathbf{A} - \lambda_k\mathbf{I}) = n - p = n - m_k$.

This lemma assures the diagonability of symmetric matrices; that is, for $\mathbf{A} = \mathbf{A}'$ the $\mathbf{U}$ in $\mathbf{AU} = \mathbf{UD}$ is nonsingular and so $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$; and as is further shown in Section 6.5.3, $\boldsymbol{U}$ is an orthogonal matrix so that for $\mathbf{A}$ being symmetric, $\mathbf{U}'\mathbf{AU} = \mathbf{D}$ and $\mathbf{A} = \mathbf{UDU}'$. Equivalent to diagonability is the fact that $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$; this means that $(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}$ has $n - (n - m_k) = m_k$ linearly independent solutions for $\mathbf{u}$; that is, the number of linearly independent eigenvectors corresponding to λ_k equals its multiplicity. ∎

7.2 OTHER RESULTS FOR SYMMETRIC MATRICES

Theorem 6.2 in Section 6.5 presented a very useful decomposition of a symmetric matrix. We have also seen in Section 6.6.1 a characterization of the eigenvalues of a symmetric matrix that is positive definite (positive semidefinite). The next section gives other interesting results concerning symmetric matrices.

7.2.1 Non-Negative Definite (n.n.d.)

The definition of a non-negative definite matrix was given in Section 5.7. Since this definition is in terms of quadratic forms, n.n.d. matrices are usually taken as being symmetric, and so also have the following properties:

(i) All eigenvalues are real (Section 6.5.1).
(ii) They are diagonable (Section 6.5.2).
(iii) Rank equals the number of nonzero eigenvalues (Section 6.5.4).

These lead, in turn, to further results, as follows.

Theorem 7.2 *The eigenvalues of a symmetric matrix are all non-negative if and only if the matrix is n.n.d.*

Proof. (Thanks go to J. C. Berry for this proof.) Let $\mathbf{A} = \mathbf{A}'$ be real. Then $\mathbf{A} = \boldsymbol{UDU}'$ for orthogonal $\boldsymbol{U}$ and $\boldsymbol{D} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ for $\lambda_i, i = 1, \ldots, n$, being the eigenvalues of $\boldsymbol{A}$. Hence, $\boldsymbol{x}'\boldsymbol{Ax} = \boldsymbol{x}'\boldsymbol{UDU}'\boldsymbol{x} = \boldsymbol{y}'\boldsymbol{Dy}$ for $\boldsymbol{y} = \boldsymbol{U}'\boldsymbol{x}$. Therefore, $\boldsymbol{A}$ is n.n.d. if, and only if, $\boldsymbol{D}$ is n.n.d. But, $\boldsymbol{D}$ is diagonal and so is n.n.d. if and only if $\lambda_i \geq 0$ for $i = 1, \ldots, n$. ∎

7.2.2 Simultaneous Diagonalization of Two Symmetric Matrices

There are at least three situations in which it is possible to find a matrix $\mathbf{P}$ that will simultaneously diagonalize two symmetric matrices (of the same order, obviously). They are as follows.

(i) $\mathbf{A}$ and $\mathbf{B}$ symmetric, with $\mathbf{AB} = \mathbf{BA}$: then there exists an orthogonal $\mathbf{P}$ such that $\mathbf{P}'\mathbf{AP} = \mathbf{D}$, a diagonal matrix of the eigenvalues of $\mathbf{A}$, and $\mathbf{P}'\mathbf{BP} = \boldsymbol{\Delta}$, some other diagonal matrix.

(ii) $\mathbf{A}$ being positive definite as well as symmetric: then there exists a $\mathbf{P}$, nonsingular but not necessarily orthogonal, such that $\mathbf{P}'\mathbf{AP} = \mathbf{I}$ and $\mathbf{P}'\mathbf{BP} = \mathbf{D}$ where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are solutions for λ to $|\mathbf{B} - \lambda\mathbf{A}| = 0$. The latter equation is sometimes called the *characteristic equation of* $\mathbf{B}$ *with respect to* $\mathbf{A}$.

(iii) $\mathbf{A}$ and $\mathbf{B}$ both non-negative definite: then there exists a nonsingular $\mathbf{P}$ such that $\mathbf{P}'\mathbf{AP}$ and $\mathbf{P}'\mathbf{BP}$ are both diagonal.

The three cases are stated as theorems and proved.

Theorem 7.3 *For symmetric* $\mathbf{A}$ *and* $\mathbf{B}$ *of the same order, there exists an orthogonal matrix* $\mathbf{P}$ *such that* $\mathbf{P}'\mathbf{AP}$ *and* $\mathbf{P}'\mathbf{BP}$ *are both diagonal if and only if* $\mathbf{AB} = \mathbf{BA}$.

Proof. This proof follows that of Graybill (1969, Theorem 12.2.12).
Sufficiency: that if $\mathbf{AB} = \mathbf{BA}$, an orthogonal $\mathbf{P}$ exists.
Because $\mathbf{A}$ is symmetric there exists $\mathbf{R}$ such that

$$\mathbf{R}'\mathbf{AR} = \mathbf{D} = \text{diag}\{\lambda_i \boldsymbol{I}_{m_i}\} \quad \text{for } \mathbf{R} \text{ orthogonal,} \tag{7.6}$$

where λ_i is one of the s distinct eigenvalues of $\mathbf{A}$ of multiplicity m_i. Let

$$\mathbf{R}'\mathbf{BR} = \mathbf{C} = \mathbf{C}' = \{\boldsymbol{C}_{ij}\} \tag{7.7}$$

where $\mathbf{C}$ is partitioned conformably with $\mathbf{D}$ in (7.6).

With $\mathbf{AB} = \mathbf{BA}$, the orthogonality of $\mathbf{R}$, and the symmetry of $\mathbf{A}$ and $\mathbf{B}$, we have $\mathbf{CD} = \mathbf{R}'\mathbf{BRR}'\mathbf{AR} = \mathbf{R}'\mathbf{BAR} = \mathbf{R}'\mathbf{ABR} = \mathbf{R}'\mathbf{ARR}'\mathbf{BR} = \mathbf{DC}$. Hence, because $\mathbf{D}$ of (7.6) is diagonal, $\lambda_j \boldsymbol{C}_{ij} = \lambda_i \boldsymbol{C}_{ij}$ for $i \neq j$ and $\lambda_i \neq \lambda_j$, and so $\boldsymbol{C}_{ij} = \mathbf{0}$ for $i \neq j$. Thus $\mathbf{C}$ of (7.7) is block diagonal:

$$\mathbf{C} = \text{diag}\{\boldsymbol{C}_{ii}\}. \tag{7.8}$$

Symmetry of $\mathbf{C}$ implies that $\boldsymbol{C}_{ii}$ is symmetric, so that there exists $\boldsymbol{Q}_i$ such that

$$\boldsymbol{Q}_i' \boldsymbol{C}_{ii} \boldsymbol{Q}_i = \Delta_{\mathbf{i}} \text{ is diagonal, with } \boldsymbol{Q}_i \text{ orthogonal.} \tag{7.9}$$

Let

$$\mathbf{Q} = \text{diag}\{\boldsymbol{Q}_i\}, \quad \text{also orthogonal,} \tag{7.10}$$

and define $\mathbf{P} = \mathbf{RQ}$, which is orthogonal because $\mathbf{R}$ and $\mathbf{Q}$ are. Then $\mathbf{P}'\mathbf{AP} = \mathbf{Q}'\mathbf{R}'\mathbf{ARQ} = \mathbf{Q}'\mathbf{DQ} = \text{diag}\{\lambda_i \boldsymbol{Q}_i'\boldsymbol{Q}_i\} = \text{diag}\{\lambda_i \boldsymbol{I}_{m_i}\} = \mathbf{D}$, using (7.6) and (7.10); and $\mathbf{P}'\mathbf{BP} = \mathbf{Q}'\mathbf{R}'\mathbf{BRQ} = \mathbf{Q}'\mathbf{CQ} = \text{diag}\{\Delta_{\mathbf{i}}\} = \Delta$, using (7.7) and (7.9). Thus with $\mathbf{P}$ being orthogonal, $\mathbf{P}'\mathbf{AP}$ and $\mathbf{P}'\mathbf{BP}$ are diagonal.

Necessity: that if $\mathbf{P}'\mathbf{AP} = \mathbf{D}$ and $\mathbf{P}'\mathbf{BP} = \Delta$ are diagonal, then $\mathbf{AB} = \mathbf{BA}$.

This is easily proved: with $\mathbf{D}$ and Δ being diagonal, $\mathbf{D}\Delta = \Delta\mathbf{D}$ and so $\mathbf{AB} = \mathbf{PP}'\mathbf{APP}'\mathbf{BPP}' = \mathbf{PD}\Delta\mathbf{P}' = \mathbf{P}\Delta\mathbf{DP}' = \mathbf{PP}'\mathbf{BPP}'\mathbf{APP}' = \mathbf{BA}$. ■

Theorem 7.4 *For symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *of the same order, with A being positive definite, there exists a nonsingular matrix* $\mathbf{P}$ *(not necessarily orthogonal), such that* $\mathbf{P}'\mathbf{AP} = \mathbf{I}$ *and* $\mathbf{P}'\mathbf{BP}$ *is a diagonal matrix of the solutions for* λ *to* $|\mathbf{B} - \lambda\mathbf{A}| = 0$.

Proof. When $\mathbf{A}$ is positive definite, it can be expressed as $\boldsymbol{A} = \boldsymbol{MM}'$ where $\boldsymbol{M}$ is nonsingular (see Corollary 6.2). Also, $\boldsymbol{M}^{-1}\boldsymbol{AM}'^{-1} = \boldsymbol{I}$. Therefore solutions $\lambda_\imath, i = 1, \dots, n$, for λ to $|\boldsymbol{M}^{-1}\boldsymbol{BM}'^{-1} - \lambda\boldsymbol{I}| = 0$ are also solutions to $|\boldsymbol{M}^{-1}||\boldsymbol{B} - \lambda\boldsymbol{A}||\boldsymbol{M}'^{-1}| = 0$, that is, to $|\mathbf{B} - \lambda\mathbf{A}| = 0$. Further, since $\boldsymbol{M}^{-1}\boldsymbol{BM}'^{-1}$ is symmetric there exists an orthogonal $\boldsymbol{Q}$ such that $\boldsymbol{Q}'\boldsymbol{M}^{-1}\boldsymbol{BM}'^{-1}\boldsymbol{Q} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Let $\boldsymbol{P} = \boldsymbol{M}'^{-1}\boldsymbol{Q}$. Then, $\boldsymbol{P}'\boldsymbol{BP} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and $\boldsymbol{P}'\boldsymbol{AP} = \boldsymbol{Q}'\boldsymbol{M}^{-1}\boldsymbol{AM}'^{-1}\boldsymbol{Q} = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$. ■

Theorem 7.5 *For real symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *of the same order and non-negative definite, there exists a nonsingular matrix* $\mathbf{P}$ *such that* $\mathbf{P}'\mathbf{AP}$ *and* $\mathbf{P}'\mathbf{BP}$ *are both diagonal.*

Proof. This proof is based on Newcombe (1960).

Let the order of $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$, with ranks a and b, respectively, and without loss of generality suppose $b \geq a$. Then, because $\mathbf{A}$ is symmetric and non-negative definite, there exists, as in 6.42, a nonsingular matrix $\mathbf{T}$ such that on defining $\gamma = n - a$,

$$\mathbf{T}'\mathbf{AT} = \mathbf{D} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix}, \tag{7.11}$$

where, for clarity, a subscripted matrix is square, the subscript being its order. Denote $\mathbf{T}'\mathbf{BT}$ by

$$\mathbf{T}'\mathbf{BT} = \begin{bmatrix} \mathbf{E}_a & \mathbf{F} \\ \mathbf{F}' & \mathbf{G}_\gamma \end{bmatrix}.$$

It is non-negative definite of rank b. Suppose $\mathbf{E}_a$ has rank $r \leq a \leq b$. Then, using the last γ rows and columns of $\mathbf{T}'\mathbf{BT}$, perform elementary operations (of the type described in Section 3.4, such as E_{ij}, $P_{ij}(\lambda)$, and $R_{ii}(\lambda)$, to be represented by $\mathbf{S}'$ and $\mathbf{S}$, respectively, so that for $\beta = b - r$ and $\delta = \gamma - \beta$ we reduce $\mathbf{T}'\mathbf{BT}$ to

$$\mathbf{S}'\mathbf{T}'\mathbf{BTS} = \begin{bmatrix} \mathbf{C}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix}, \tag{7.12}$$

where $\mathbf{C}_a$ has rank r. The operations are to be done in such a way that the first a rows (columns) of $\mathbf{T}'\mathbf{B}\mathbf{T}$ are involved only by adding multiples of the other γ rows (columns) to them. Therefore, performing the operations represented by $\mathbf{S}$ on $\mathbf{T}'\mathbf{A}\mathbf{T}$ of (7.11) will have no effect on (7.11), since all the last γ rows and columns of (7.11) are null, that is,

$$\mathbf{S}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{S} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix}. \tag{7.13}$$

Since $\mathbf{C}_a$ of (7.12) is symmetric of rank r, there is a $\mathbf{Q}_a$ such that

$$\begin{aligned} \mathbf{Q}'_a\mathbf{C}_a\mathbf{Q}_a &= \mathbf{D}_a \\ &= \boldsymbol{D}\{\lambda_1 \ldots \lambda_p 0 \ldots 0\} \quad \text{for } \boldsymbol{Q}_a \text{ orthogonal.} \end{aligned} \tag{7.14}$$

Therefore, on defining

$$\mathbf{R} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_\delta \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\gamma \end{bmatrix} \tag{7.15}$$

we have from (7.12), (7.14), and (7.15)

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{S}\mathbf{R} = \begin{bmatrix} \mathbf{D}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix} = \Delta, \quad \text{diagonal;}$$

and from (7.13) and (7.15) we have

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{S}\mathbf{R} = \begin{bmatrix} \mathbf{Q}'_a\mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \mathbf{D}, \quad \text{diagonal.}$$

Hence, for $\mathbf{P} = \mathbf{T}\mathbf{S}\mathbf{R}$ we have $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{D}$ and $\mathbf{P}'\mathbf{B}\mathbf{P} = \Delta$, both diagonal. ∎

Theorem 7.3 can be extended to include the simultaneous diagonalization of several symmetric matrices. This extension is given in the following theorem whose proof is similar to that of Theorem 7.3:

Theorem 7.6 *Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k$ be symmetric matrices of the same order. Then there exists an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{A}_1\boldsymbol{P}', \boldsymbol{P}\boldsymbol{A}_2\boldsymbol{P}', \ldots, \boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}'$ are all diagonal if and only if $\boldsymbol{A}_i\boldsymbol{A}_j = \boldsymbol{A}_j\boldsymbol{A}_i$ for all $i, j = 1, 2, \ldots, k$.*

Theorem 7.6 has interesting applications in linear models, as will be seen later.

7.3 THE CAYLEY–HAMILTON THEOREM

Theorem 7.7 *Let* $\boldsymbol{A}$ *be a matrix of order* $n \times n$ *with the characteristic equation* $|\boldsymbol{A} - \lambda \boldsymbol{I}_n| = 0$, *which can be written as*

$$|\boldsymbol{A} - \lambda \boldsymbol{I}_n| = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n$$
$$= 0, \tag{7.16}$$

where $a_0, a_1, \ldots, a_n$ *are known coefficients that depend on the elements of* $\boldsymbol{A}$. *Then,* $\boldsymbol{A}$ *satisfies its characteristic equation, that is,*

$$a_0\boldsymbol{I}_n + a_1\boldsymbol{A} + a_2\boldsymbol{A}^2 + \cdots + a_n\boldsymbol{A}^n = \boldsymbol{0}. \tag{7.17}$$

Proof. Let $\boldsymbol{B} = \mathrm{adj}(\boldsymbol{A} - \lambda \boldsymbol{I}_n)$ be the adjugate, or adjoint, matrix of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ (see Section 4.12 for a definition of the adjugate matrix), where λ is not an eigenvalue of $\boldsymbol{A}$. By the definition of the adjugate we have

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)^{-1} = \frac{adj(\boldsymbol{A} - \lambda \boldsymbol{I}_n)}{|\boldsymbol{A} - \lambda \boldsymbol{I}_n|}. \tag{7.18}$$

Since the elements of $\boldsymbol{B}$ are polynomials of degree $n-1$ or less in λ, we can express $\boldsymbol{B}$ as

$$\boldsymbol{B} = \boldsymbol{B}_0 + \lambda\boldsymbol{B}_1 + \lambda^2\boldsymbol{B}_2 + \cdots + \lambda^{n-1}B_{n-1}, \tag{7.19}$$

where $\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{n-1}$ are matrices of order $n \times n$ that do not depend on λ. Using (7.18) we can write

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{B} = |\boldsymbol{A} - \lambda \boldsymbol{I}_n|\boldsymbol{I}_n$$
$$(a_0 + a_1\lambda + \cdots + a_n\lambda^n)\boldsymbol{I}_n. \tag{7.20}$$

Substituting the representation of $\boldsymbol{B}$ given by (7.19) in (7.20) we get

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)(\boldsymbol{B}_0 + \lambda\boldsymbol{B}_1 + \cdots + \lambda^{n-1}\boldsymbol{B}_{n-1}) = (a_0 + a_1\lambda + \cdots + a_n\lambda^n)\boldsymbol{I}_n. \tag{7.21}$$

By comparing the coefficients of the powers of λ on both sides of (7.21), we obtain

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{B}_0 &= a_0\boldsymbol{I}_n \\
\boldsymbol{A}\boldsymbol{B}_1 - \boldsymbol{B}_0 &= a_1\boldsymbol{I}_n \\
\boldsymbol{A}\boldsymbol{B}_2 - \boldsymbol{B}_1 &= a_2\boldsymbol{I}_n \\
&\vdots \\
\boldsymbol{A}\boldsymbol{B}_{n-1} - \boldsymbol{B}_{n-2} &= a_{n-1}\boldsymbol{I}_n \\
-\boldsymbol{B}_{n-1} &= a_n\boldsymbol{I}_n.
\end{aligned}$$

Multiplying on the left the second equality by $\boldsymbol{A}$, the third by $\boldsymbol{A}^2$, etc, the one before last by $\boldsymbol{A}^{n-1}$, and the last one by $\boldsymbol{A}^n$, then adding up all the resulting equalities, we finally conclude that (7.17) is true for any square matrix. This indicates that the matrix $\boldsymbol{A}$ satisfies its own characterisation equation. ■

A useful application of Cayley–Hamilton's theorem is that when the characteristic equation is known, even if the eigenvalues are not, the nth and successive powers of $\boldsymbol{A}$ can be obtained as polynomials of $\boldsymbol{A}$. This is also true for the inverse of $\boldsymbol{A}$, if $\boldsymbol{A}$ is nonsingular.

Example 7.2 *For*

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix}, \quad \mathbf{A}^2 = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{14}\begin{bmatrix} 5 & -3 \\ -2 & 4 \end{bmatrix}$$

and the characteristic equation of $\boldsymbol{A}$ *is*

$$\lambda^2 - 9\lambda + 14 = 0.$$

The Cayley–Hamilton theorem is satisfied because

$$\mathbf{A}^2 - 9A + 14\mathbf{I} = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix} - 9\begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix} + 14\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Polynomials for obtaining powers of $\mathbf{A}$ *of degree 2 or more come from rewriting* $\mathbf{A}^2 - 9\mathbf{A} + 14\mathbf{I} = 0$ *as* $\mathbf{A}^2 = 9\mathbf{A} - 14\mathbf{I}$. *Hence,*

$$\mathbf{A}^3 = 9\mathbf{A}^2 - 14\mathbf{A} = 9\,(9\mathbf{A} - 14\mathbf{I}) - 14\mathbf{A} = 67\mathbf{A} - 126\mathbf{I}$$

and

$$\mathbf{A}^4 = 67\mathbf{A}^2 - 126\mathbf{A} = 67\,(9\mathbf{A} - 14\mathbf{I}) - 126\mathbf{A} = 477\mathbf{A} - 938\mathbf{I}.$$

This means that recurrence relations can be established between the coefficients in successive powers of $\mathbf{A}$, *so enabling* $\mathbf{A}^k$, *for* $k \geq n$, *to be written as a polynomial in* $\mathbf{A}$ *of degree* $n - 1$, *the coefficients being functions of* k. *Thus for the above example it can be shown that, for* $k \geq 2$,

$$\mathbf{A}^k = \left(\frac{1}{5}\right)\left(7^k - 2^k\right)\mathbf{A} - \left(\frac{14}{5}\right)\left(7^{k-1} - 2^{k-1}\right)\mathbf{I}.$$

7.4 THE SINGULAR-VALUE DECOMPOSITION

Theorem 7.8 *Let* $\boldsymbol{A}$ *be an* $m \times n$ *matrix* $(m \leq n)$ *of rank* r. *There exist orthogonal matrices* $\boldsymbol{P}$, *of order* $m \times m$, *and* $\boldsymbol{Q}$, *of order* $n \times n$, *such that*

$$\boldsymbol{A} = \boldsymbol{P}[\boldsymbol{D} \quad \mathbf{0}]\boldsymbol{Q}', \tag{7.22}$$

where $\boldsymbol{D}$ *is an* $m \times m$ *diagonal matrix with non-negative diagonal elements* d_i, $i = 1, 2, \ldots, m$, *and* $\mathbf{0}$ *is a zero matrix of order* $m \times (n - m)$. *The positive diagonal elements* *of* $\boldsymbol{D}$ *are the square roots of the positive eigenvalues of* $\boldsymbol{AA}'$ *(or, equivalently, of* $\boldsymbol{A}'\boldsymbol{A}$*), and are called the singular values of* $\boldsymbol{A}$*. If* $m = n$*, then* $\boldsymbol{P}'\boldsymbol{AQ} = \boldsymbol{D}$*.*

Proof. Assume that $m < n$. The matrix $\boldsymbol{AA}'$ is symmetric, positive semidefinite of rank r. Then, there exists an orthogonal matrix $\boldsymbol{P}$ such that

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{AA}'\boldsymbol{P} &= \boldsymbol{D}^2\\ &= \operatorname{diag}\left(d_1^2, d_2^2, \ldots, d_m^2\right)\\ &= \operatorname{diag}\left(d_1^2, d_2^2, \ldots, d_r^2, 0, 0, \ldots, 0\right)\end{aligned}$$

where $d_i^2 \neq 0$, $i = 1, 2, \ldots, r$. Let us partition $\boldsymbol{P}$ as

$$\boldsymbol{P} = \begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2\end{bmatrix},$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are of orders $m \times r$ and $m \times (m - r)$, respectively. We then have

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{AA}'\boldsymbol{P} &= \begin{bmatrix}\boldsymbol{P}_1' \\ \boldsymbol{P}_2'\end{bmatrix}\boldsymbol{AA}'\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2\end{bmatrix}\\ &= \operatorname{diag}(\boldsymbol{D}_1^2, \boldsymbol{0}),\end{aligned}$$

where $\boldsymbol{D}_1^2 = \operatorname{diag}(d_1^1, d_2^2, \ldots, d_r^2)$. It follows that

$$\boldsymbol{P}_1'\boldsymbol{AA}'\boldsymbol{P}_1 = \boldsymbol{D}_1^2, \text{ or } (\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{A})(\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{A})' = \boldsymbol{I}_r, \tag{7.23}$$

and

$$\boldsymbol{P}_2'\boldsymbol{AA}'\boldsymbol{P}_2 = \boldsymbol{0}, \text{ or } (\boldsymbol{P}_2'\boldsymbol{A})(\boldsymbol{P}_2'\boldsymbol{A})' = \boldsymbol{0}.$$

Hence, $\boldsymbol{P}_2'\boldsymbol{A} = \boldsymbol{0}$. Let the matrix $\boldsymbol{Q}_1$ be defined as

$$\boldsymbol{Q}_1 = \boldsymbol{A}'\boldsymbol{P}_1\boldsymbol{D}_1^{-1}. \tag{7.24}$$

This matrix is of order $n \times r$ whose columns are orthonormal since $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_r$ by (7.23). Furthermore, there exits a matrix $\boldsymbol{Q}_2$ of order $n \times (n - r)$ whose columns are orthonormal and are orthogonal to the columns of $\boldsymbol{Q}_1$. Then, the matrix

$$\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix}$$

is orthogonal of order $n \times n$. We then have

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{AQ} &= \begin{bmatrix}\boldsymbol{P}_1' \\ \boldsymbol{P}_2'\end{bmatrix}\boldsymbol{A}\begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{P}_1'\boldsymbol{AQ}_1 & \boldsymbol{P}_1'\boldsymbol{AQ}_2 \\ \boldsymbol{P}_2'\boldsymbol{AQ}_1 & \boldsymbol{P}_2'\boldsymbol{AQ}_2\end{bmatrix}.\end{aligned} \tag{7.25}$$

Note that $\boldsymbol{P}_2'\boldsymbol{A}\boldsymbol{Q}_1$ and $\boldsymbol{P}_2'\boldsymbol{A}\boldsymbol{Q}_2$ are each equal to a zero matrix since $\boldsymbol{P}_2'\boldsymbol{A} = \boldsymbol{0}$. Furthermore, $\boldsymbol{P}_1'\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{0}$ since by the definition of the matrix $\boldsymbol{Q}$,

$$\boldsymbol{Q}_1'\boldsymbol{Q}_2 = \boldsymbol{0}. \tag{7.26}$$

Using the expression for $\boldsymbol{Q}_1$ given by (7.24) in (7.26) we get, $\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{0}$. This implies that $\boldsymbol{P}_1'\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{0}$. Making the substitution in (7.25), we get

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{A}\boldsymbol{Q} &= \begin{bmatrix} \boldsymbol{P}_1'\boldsymbol{A}\boldsymbol{A}'\boldsymbol{P}_1\boldsymbol{D}_1^{-1} & \boldsymbol{0}_{r\times(n-r)} \\ \boldsymbol{0}_{(m-r)\times r} & \boldsymbol{0}_{(m-r)\times(n-r)} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0}_{r\times(n-r)} \\ \boldsymbol{0}_{(m-r)\times r} & \boldsymbol{0}_{(m-r)\times(n-r)} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0}_{m\times(n-m)} \end{bmatrix},\end{aligned} \tag{7.27}$$

since $n - r = (m - r) + (n - m)$ and the fact that $\boldsymbol{P}_1'\boldsymbol{A}\boldsymbol{A}'\boldsymbol{P}_1 = \boldsymbol{D}_1^2$ as shown in (7.23). We finally conclude that

$$\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D} & \boldsymbol{0}_{m\times(n-m)} \end{bmatrix}\boldsymbol{Q}'.$$

■

Corollary 7.2 *The columns of the matrix Q in Theorem 7.8 are orthonormal eigenvectors of* $\boldsymbol{A}'\boldsymbol{A}$ *which has the same nonzero eigenvalues as those of* $\boldsymbol{A}\boldsymbol{A}'$.

Proof. Using (7.27) we have

$$\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0}_{r\times(n-r)} \\ \boldsymbol{0}_{(m-r)\times r} & \boldsymbol{0}_{(m-r)\times(n-r)} \end{bmatrix}\boldsymbol{Q}'. \tag{7.28}$$

Transposing both sides of (7.28) we get

$$\boldsymbol{A}' = \boldsymbol{Q}\begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0}_{r\times(m-r)} \\ \boldsymbol{0}_{(n-r)\times r} & \boldsymbol{0}_{(}n-r)\times(m-r) \end{bmatrix}\boldsymbol{P}'. \tag{7.29}$$

Using (7.28) and (7.29) and noting that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_m$ we conclude

$$\begin{aligned}\boldsymbol{A}'\boldsymbol{A} &= \boldsymbol{Q}\begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0}_{r\times(m-r)} \\ \boldsymbol{0}_{(n-r)\times r} & \boldsymbol{0}_{(n-r)\times(m-r)} \end{bmatrix}\begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0}_{r\times(n-r)} \\ \boldsymbol{0}_{(m-r)\times r} & \boldsymbol{0}_{(m-r)\times(n-r)} \end{bmatrix}\boldsymbol{Q}' \\ &= \boldsymbol{Q}\begin{bmatrix} \boldsymbol{D}_1^2 & \boldsymbol{0}_{r\times(n-r)} \\ \boldsymbol{0}_{(n-r)\times r} & \boldsymbol{0}_{(n-r)\times(n-r)} \end{bmatrix}\boldsymbol{Q}'.\end{aligned}$$

This clearly indicates that $\boldsymbol{Q}$ is an orthogonal matrix of eigenvectors of $\boldsymbol{A}'\boldsymbol{A}$ with nonzero eigenvalues given by the diagonal elements of $\boldsymbol{D}_1^2$. ■

Example 7.3 *For*

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{AA}' = \begin{bmatrix} 6 & 2 & 4 \\ 2 & 6 & 4 \\ 4 & 4 & 4 \end{bmatrix}$$

has nonzero eigenvalues 12 and 4. The rank of $\mathbf{A}$ *is therefore equal to 2. The matrix* $\boldsymbol{P}$ *of orthonormal eigenvectors of* $\mathbf{AA}'$ *is given by*

$$\boldsymbol{P} = \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix} / \sqrt{6}.$$

The matrix $\boldsymbol{A}'\boldsymbol{A}$ *is*

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 5 & 1 & 3 & 3 \\ 1 & 5 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

Its matrix $\boldsymbol{Q}$ *of orthonormal eigenvectors is*

$$\boldsymbol{Q} = \begin{bmatrix} \sqrt{3} & \sqrt{6} & \sqrt{2} & 1 \\ \sqrt{3} & -\sqrt{6} & \sqrt{2} & 1 \\ \sqrt{3} & 0 & -2\sqrt{2} & 1 \\ \sqrt{3} & 0 & 0 & -3 \end{bmatrix} / \sqrt{12}.$$

In addition, the matrix $\boldsymbol{D}$ *is*

$$\boldsymbol{D} = diag(\sqrt{12}, 2, 0).$$

Applying formula (7.22) we get

$$\boldsymbol{A} = \boldsymbol{P}[\boldsymbol{D} \quad \mathbf{0}_{3\times 1}]\boldsymbol{Q}'. \tag{7.30}$$

The reader can verify that the product of the three matrices on the right-hand side of (7.30) is equal to $\boldsymbol{A}$.

Good (1969) described several interesting applications of the singular-values decomposition in statistics, including the theory of least squares and the analysis of contingency tables.

REFERENCES

Banerjee, S. and Roy, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, Boca Raton, FL.

Good, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics*, 11, 823–831.

Graybill, F. A. (1969). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth International Group, Belmont, CA.

Newcombe, R. W. (1960). On the simultaneous diagonalization of two semidefinite matrices. *Quart. Appl. Math.*, 19, 144–146.

EXERCISES

7.1 Show that if $\boldsymbol{A}$ is an idempotent matrix, then rank($\boldsymbol{A}$) = $tr(\boldsymbol{A})$.

7.2 Find the singular-value decomposition of $\begin{bmatrix} 10 & -5 \\ 2 & -11 \\ 6 & -8 \end{bmatrix}$.

7.3 Find the rank of $\mathbf{M} = \begin{bmatrix} 1 & -3 & 5 \\ 0 & 1 & 7 \\ 2 & -4 & 28 \\ 3 & -13 & 0 \\ 4 & -14 & -3 \\ 5 & -9 & 30 \\ 6 & 2 & 173 \end{bmatrix}$.

7.4 Suppose

$$\mathbf{K} = \mathbf{K}', \quad \mathbf{K} = \mathbf{K}^3, \quad \mathbf{K1} = \mathbf{0}, \quad \text{and} \quad \mathbf{K}\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}.$$

Calculate the following values, in each case giving reasons as to why the value can be calculated *without calculating* **K**.

(a) The order of **K**.

(b) The rank of **K**.

(c) The trace of **K**.

(d) The determinant of **K**.

7.5 Let $\boldsymbol{A}$ be a matrix of order $n \times n$ that satisfies the equation

$$\boldsymbol{A}^2 + 2\boldsymbol{A} + \boldsymbol{I}_n = \mathbf{0}.$$

(a) Show that $\boldsymbol{A}$ is nonsingular.

(b) Find the inverse of $\boldsymbol{A}$.

7.6 Suppose that there is an $n \times n$ matrix $\boldsymbol{A} \neq \boldsymbol{0}$ such that $\boldsymbol{A}^n = \boldsymbol{0}$. Show that such a matrix is not diagonable [see Banerjee and Roy (2014), Section 11.4].

7.7 Find a 2×2 matrix that is not diagonable.

7.8 Suppose that $\boldsymbol{A}$ is an $n \times n$ matrix and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where $\boldsymbol{P}$ is a nonsingular matrix. Find such a $\boldsymbol{P}$ matrix and the diagonal elements of $\boldsymbol{\Lambda}$ [see Banerjee and Roy (2014), Section 11.4].

7.9 Find a matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is diagonal, where

$$\boldsymbol{A} = \begin{bmatrix} 6 & 3 & 2 \\ -5 & 0 & 3 \\ 7 & 5 & 1 \end{bmatrix}.$$

7.10 Show that the matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \begin{bmatrix} 5 & 3 \\ 0 & 5 \end{bmatrix}$$

is not diagonable.

8

Generalized Inverses

This chapter concerns the development for rectangular and singular matrices of an analogue of the inverse of nonsingular matrices. The resulting matrices, known generically as generalized inverses, play an important role in understanding the solutions to linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ when $\boldsymbol{A}$ has no inverse. With this in mind, because such equations arise in many important statistical applications (e.g., regression analysis and linear models), we give here a brief account of generalized inverses and of some of their properties, especially as they pertain to solving linear equations. The reader wishing to pursue the topic in more detail is referred to books by Albert (1972), Ben-Israel and Greville (1974), Nashed and Zuhair (1976), Pringle and Rayner (1971), Rao and Mitra (1971), Graybill (1983), Harville (1997), and Gruber (2014).

8.1 THE MOORE–PENROSE INVERSE

Given any matrix $\mathbf{A}$, there is a unique matrix $\mathbf{M}$ such that

$$\begin{aligned} &\text{(i) } \mathbf{AMA} = \mathbf{A} \qquad &\text{(iii) } \mathbf{AM} \text{ is symmetric} \\ &\text{(ii) } \mathbf{MAM} = \mathbf{M} \qquad &\text{(iv) } \mathbf{MA} \text{ is symmetric.} \end{aligned} \tag{8.1}$$

This result is developed in Penrose (1955), on foundations laid by Moore (1920). The Penrose paper established not only the existence of $\mathbf{M}$ but also its uniqueness for a given $\mathbf{A}$. One way of writing $\mathbf{M}$ is based on the factoring of $\mathbf{A}_{p\times q}$ as $\mathbf{A} = \mathbf{KL}$ as in Section 4.18,

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

where $\mathbf{K}$ and $\mathbf{L}$ have full column and row rank, respectively, equal to $r_{\mathbf{A}}$. Then $\mathbf{M}$ of (8.1) is

$$\mathbf{M} = \mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}'. \tag{8.2}$$

Readers should verify for themselves that the expression in (8.2) does satisfy (8.1), using the nonsingularity of $\mathbf{K}'\mathbf{K}$ and $\mathbf{L}\mathbf{L}'$ to do so.

Example 8.1 *The factoring given by the theorem of Section (4.18), namely* $\mathbf{A} = \mathbf{KL}$, *is exemplified by*

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} = \mathbf{KL}. \tag{8.3}$$

From this,

$$\boldsymbol{K}'\boldsymbol{A}\boldsymbol{L}' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} = 6 \begin{bmatrix} 1 & -2 \\ -2 & 10 \end{bmatrix}.$$

Hence from (8.2)

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \left(\frac{1}{36}\right) \begin{bmatrix} 10 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 4 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}. \tag{8.4}$$

It is easily verified that $\mathbf{A}$ *and* $\boldsymbol{M}$ *satisfy the four equations in (8.1).*

$\boldsymbol{M}$ *is generally known as the Moore–Penrose inverse of* $\boldsymbol{A}$*; and the four conditions in (8.1) are usually called the Penrose conditions. For* $\mathbf{A}$ *of order* $p \times q$*, the order of* $\boldsymbol{M}$ *is* $q \times p$*. When* $\mathbf{A}$ *is nonsingular,* $\boldsymbol{K}$ *and* $\mathbf{L}$ *are likewise and the Moore–Penrose inverse is the regular inverse,* $\boldsymbol{M} = \boldsymbol{A}^{-1}$.

8.2 GENERALIZED INVERSES

The matrix $\mathbf{M}$ defined by the four Penrose conditions in (8.1) is unique for a given $\mathbf{A}$. But there are many matrices $\mathbf{G}$ which satisfy just the first Penrose condition:

$$\mathbf{AGA} = \mathbf{A}. \tag{8.5}$$

Nevertheless, they are of such importance in solving linear equations that we direct most attention to those matrices $\mathbf{G}$ rather than to the Moore–Penrose inverse $\mathbf{M}$.

Any matrix $\mathbf{G}$ satisfying (8.5) is called a generalized inverse of $\mathbf{A}$; and, by (8.5), when $\mathbf{A}$ is $p \times q$ then $\mathbf{G}$ is $q \times p$. Although the name generalized inverse has not been adopted universally, it is widely used; other names are discussed in Section 8.3. Notice that $\mathbf{G}$ is "a" generalized inverse of $\mathbf{A}$ and not "the" generalized inverse, because for any given $\mathbf{A}$ there are generally many matrices $\mathbf{G}$ satisfying (8.5). The exception is when $\mathbf{A}$ is nonsingular, in which case there is only one $\mathbf{G}$ satisfying (8.5) and it is the regular inverse $\mathbf{G} = \mathbf{A}^{-1} = \mathbf{M}$.

Generalized inverses are associated with systems of linear equations that are consistent.

Definition 8.1 *The linear equations,*

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \tag{8.6}$$

are consistent if $\mathbf{y}$ *belongs to the column space of* $\boldsymbol{A}$.

For example, the equations

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \end{bmatrix},$$

are consistent. If, however, the second element of $\boldsymbol{y}$ in the above equations is changed from 21 to 24, then it is easy to verify that the above equations are no longer consistent since they cannot be solved for x_1 and $\boldsymbol{x}_2$.

If the linear equations in (8.6) are consistent, then $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{y}$ is a solution to these equations. Two derivations of $\boldsymbol{G}$ are now given.

8.2.1 Derivation Using the Singular-Value Decomposition

Let $\boldsymbol{A}$ be a matrix of order $m \times n$ $(m \leq n)$ of rank r. By Theorem 7.8 there exist orthogonal matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of orders $m \times m$ and $n \times n$, respectively such that

$$\boldsymbol{A} = \boldsymbol{P} \begin{bmatrix} \boldsymbol{D} & \mathbf{0}_{m \times (n-m)} \end{bmatrix} \boldsymbol{Q}', \tag{8.7}$$

where $\boldsymbol{D} = \text{diag}(d_1, d_2, \ldots, d_r, 0, 0, \ldots, 0)$; $d_1, d_2, \ldots, d_r$ are the positive square roots of the positive eigenvalues of $\boldsymbol{A}\boldsymbol{A}'$ (or of $\boldsymbol{A}'\boldsymbol{A}$) which are followed by $m - r$ zero eigenvalues of $\boldsymbol{A}\boldsymbol{A}'$. Then, a generalized inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{G} = \boldsymbol{Q}\boldsymbol{G}^*\boldsymbol{P}', \tag{8.8}$$

where

$$\boldsymbol{G}^* = \begin{bmatrix} \boldsymbol{D}^* \\ \mathbf{0}_{(n-m) \times m} \end{bmatrix}, \tag{8.9}$$

and $\boldsymbol{D}^*$ is a diagonal matrix of order $m \times m$ of the form

$$\boldsymbol{D}^* = \text{diag}\left(d_1^{-1}, d_2^{-1}, \ldots, d_r^{-1}, 0, 0, \ldots, 0\right). \tag{8.10}$$

The number of zeros in (8.10) is $m - r$. Using (8.7) and (8.8), it can be verified that

$$\boldsymbol{AGA} = \boldsymbol{A}. \tag{8.11}$$

Note that $\boldsymbol{G}^*$ is of order $n \times m$.

8.2.2 Derivation Based on Knowing the Rank

Suppose that in $\mathbf{A}_{m\times n}$ the leading principal submatrix $\mathbf{A}_{11}$ is nonsingular of rank r. Then a generalized inverse obtainable from partitioning $\mathbf{A}$ as

$$\mathbf{A}_{m\times n} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \text{ is } \quad \mathbf{G}_{n\times m} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{8.12}$$

where the null matrices in $\mathbf{G}$ have appropriate order to make $\mathbf{G}$ be of order $n \times m$. The verification of $\mathbf{AGA} = \mathbf{A}$ is done as follows:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{12} & \boldsymbol{A}_{22} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix}. \end{aligned} \tag{8.13}$$

Since the matrix $\boldsymbol{A}$ is of rank r and its first r rows are linearly independent, the portion $[\boldsymbol{A}_{21} \quad \boldsymbol{A}_{22}]$ of $\boldsymbol{A}$ must have its rows linearly dependent on the rows of $[\boldsymbol{A}_{11} \quad \boldsymbol{A}_{12}]$. Hence,

$$[\boldsymbol{A}_{21} \quad \boldsymbol{A}_{22}] = \boldsymbol{K}[\boldsymbol{A}_{11} \quad \boldsymbol{A}_{12}],$$

for some matrix $\boldsymbol{K}$. It follows that $\boldsymbol{K} = \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$ and $\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} = \boldsymbol{A}_{22}$. Making the substitution in (8.13), we conclude that $\boldsymbol{AGA} = \boldsymbol{A}$. Therefore, $\boldsymbol{G}$ is a generalized inverse of $\boldsymbol{A}$.

Example 8.2 *A generalized inverse of*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} \textit{ is } \quad \mathbf{G} = \frac{1}{7}\begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A General Form There is no need for the nonsingular submatrix of order $r \times r$ to be in the leading position. It can be anywhere in $\boldsymbol{A}$, whereupon the following algorithm can be developed.

Suppose the nonsingular submatrix is not in the leading position. Let **R** and **S** be permutation matrices (Section 4.17) such that **RAS** brings a nonsingular submatrix of order $r \times r$ to the leading position; that is,

$$\mathbf{RAS} = \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \tag{8.14}$$

with $\mathbf{B_{11}}$ being nonsingular of order $r \times r$, Then, a generalized inverse of **B** is

$$\mathbf{F} = \begin{bmatrix} \mathbf{B}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{and} \quad \mathbf{G} = \mathbf{SFR} \tag{8.15}$$

is a generalized inverse of **A**. Now, because of the orthogonality of the permutation matrices, (8.14) implies

$$\mathbf{A} = \mathbf{R}'\mathbf{B}\mathbf{S}' = \mathbf{R}' \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \mathrm{S}'. \tag{8.16}$$

Insofar as $\mathbf{B_{11}}$ is concerned, this product represents the operations of returning the elements of $\mathbf{B_{11}}$ to their original positions in **A**. Hence, because (8.15) is equivalent to

$$\mathbf{G} = \mathbf{SFR} = (\mathbf{R}'\mathbf{F}'\mathbf{S}')' = \left\{ \mathbf{R}' \begin{bmatrix} (\mathbf{B}_{11}^{-1})' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{S}' \right\}', \tag{8.17}$$

we see by comparing (8.17) with (8.16) that the product involving $\mathbf{R}'$ and $\mathbf{S}'$ in (8.17) represents putting the elements of $(\mathbf{B}_{11}^{-1})'$ into the corresponding positions (of $\mathbf{G}'$) that the elements of $\mathbf{B}_{11}$ occupied in **A**. Hence, (8.17) gives the following algorithm for finding **G**.

(i) In **A** find any nonsingular submatrix of order $r \times r$. Denote it by **W** [in place of $\mathbf{B}_{11}$ in (8.14)]. It is not necessary for **W** to come from r adjacent rows and columns of $\boldsymbol{A}$.

(ii) Invert and transpose $\mathbf{W} : (\mathbf{W}^{-1})'$.

(iii) In **A** replace each element of **W** by the corresponding element of $(\mathbf{W}^{-1})'$; that is, when $a_{ij} = w_{st}$, the (s, t)th element of **W**, then replace a_{ij} by $w^{t,s}$, the (t, s)th element of $\mathbf{W}^{-1}$, equivalent to the (s, t)th element of the transpose of $\mathbf{W}^{-1}$.

(iv) Replace all other elements of **A** by zero.

(v) Transpose the resulting matrix.

(vi) The result is **G**, a generalized inverse of **A**.

Note that this procedure is *not* equivalent, in (iii), to replacing elements of **W** in **A** by the elements of $\mathbf{W}^{-1}$ (and others by zero) and then in (v) transposing. It is if **W** is symmetric. Nor is it equivalent to replacing, in (iii), elements of **W** in **A** by elements of $\mathbf{W}^{-1}$ (and others by zero) and then in (v) not transposing. In general, the algorithm must be carried out exactly as described.

Example 8.3 *The matrix*

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

has the following matrices, among others, as generalized inverses:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1.5 & 2.5 \\ 0 & .5 & -.5 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & -.5 & 1.5 \\ 0 & 0 & 0 \\ 0 & .1 & -.1 \end{bmatrix} \text{ and } \begin{bmatrix} .25 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -.15 & 0 & .2 \end{bmatrix}.$$

They are derived from inverting the 2×2 *submatrices*

$$\begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 15 \\ 1 & 5 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}, \quad \text{respectively.}$$

The algorithm can be simplified slightly for symmetric matrices as discussed in Section 8.4.

8.3 OTHER NAMES AND SYMBOLS

A useful alternative symbol for $\mathbf{G}$ satisfying $\mathbf{AGA} = \mathbf{A}$ is $\mathbf{A}^-$;

$$\mathbf{G} \equiv \mathbf{A}^-. \tag{8.18}$$

A *reflexive generalized inverse*, to be denoted by $\mathbf{A}_r^-$, is a matrix satisfying the first two Penrose conditions:

$$\mathbf{A}\mathbf{A}_r^-\mathbf{A} = \mathbf{A} \quad \text{and} \quad \mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^- = \mathbf{A}_r^-. \tag{8.19}$$

Given a generalized inverse $\mathbf{A}^-$ it is easily shown that

$$\mathbf{A}_r^- = \mathbf{A}^-\mathbf{A}\mathbf{A}^- \tag{8.20}$$

is a reflexive generalized inverse satisfying (8.19).

The literature abounds with names for these inverses: pseudo-inverse, p-inverse, g-inverse, effective inverse, normalized generalized inverse and weak generalized inverse, and others (e.g., Searle 1971, p. 19). Any reader encountering these needs to be sure of understanding the meanings given to them, because they are not universal. Some, for example, refer to a reflexive generalized inverse that also satisfies one but not both of the Penrose symmetry conditions, namely (iii) or (iv) in (8.1).

Notation also comes in many forms, one popular variant being to use subscripts for denoting which of the Penrose conditions are satisfied. For example, a g_1-inverse would be $\mathbf{A}^-$ or $\mathbf{G}$, satisfying just condition (i), $\mathbf{AGA} = \mathbf{A}$; similarly, a g_{12}-inverse would be a reflexive generalized inverse and the g_{1234}-inverse would be the Moore-Penrose inverse. This notation is useful when dealing with matrices satisfying different combinations of the Penrose conditions. In our case, when focusing prime attention on those satisfying just the first condition, it is easier to use a single symbol $\mathbf{G}$ defined by $\mathbf{AGA} = \mathbf{A}$, although the alternative $\mathbf{A}^-$ is frequently used.

Theorem 8.1 *Let $\boldsymbol{G}$ be a generalized inverse of $\boldsymbol{A}$. Then,*

(i) *$\boldsymbol{GA}$ is idempotent.*
(ii) *$r(A) = r(GA)$.*

Proof.

(i) This follows from $\boldsymbol{GAGA} = \boldsymbol{GA}$.
(ii) By property (xiv) in Section 4.15, $r(\boldsymbol{GA}) \leq r(\boldsymbol{A})$. But, $\boldsymbol{A} = \boldsymbol{AGA}$. By applying again the same property, we get $r(\boldsymbol{A}) \leq r(\boldsymbol{GA})$. We conclude that $r(\boldsymbol{A}) = r(\boldsymbol{GA})$.

■

8.4 SYMMETRIC MATRICES

Generalized inverses of symmetric matrices have some interesting properties. First, from the diagonal form of $\mathbf{A} = \mathbf{A}'$, obtained on the basis of the spectral decomposition theorem (Theorem 6.2), we obtain

$$\boldsymbol{A} = \boldsymbol{P\Lambda P}',$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$ along its diagonal and $\boldsymbol{P}$ is an orthogonal matrix whose columns are the corresponding eigenvectors of $\boldsymbol{A}$. If $\boldsymbol{A}$ is of order $n \times n$ and rank r, the first r diagonal elements of $\boldsymbol{\Lambda}$ can be chosen to be the nonzero eigenvalues of $\boldsymbol{A}$ and the remaining $n - r$ elements are the zero eigenvalues. Then, a generalized inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{G} = \boldsymbol{P} \begin{bmatrix} \boldsymbol{D}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{P}', \tag{8.21}$$

where $\boldsymbol{D}_r$ is the diagonal matrix of nonzero eigenvalues of $\boldsymbol{A}$. It can be verified that $\boldsymbol{A} = \boldsymbol{AGA}$. Transposing both sides of this equation gives

$$\boldsymbol{A} = \boldsymbol{AG}'\boldsymbol{A}. \tag{8.22}$$

This indicates that if $\boldsymbol{G}$ is a generalized inverse of a symmetric matrix $\boldsymbol{A}$, then so is the transpose $\boldsymbol{G}'$, whether $\boldsymbol{G}$ is symmetric or not.

8.4.1 A General Algorithm

The general form of the algorithm in Section 8.2.2 can be simplified when A is symmetric. Principal submatrices of **A** are then symmetric and so by using them, the transposing in (iii) and (v) can be ignored. The algorithm can then become as follows.

(i) In **A** find a nonsingular principal submatrix, **W**, of order $r \times r$.
(ii) Invert **W**.
(iii) In **A** replace each element w_{ij} of **W** by the element w^{ij} of $\mathbf{W}^{-1}$.
(iv) Replace all other elements of **A** by zero.
(v) The result is **G**, a generalized inverse of **A**.

Of course, when **A** is symmetric and a nonsymmetric, nonprincipal submatrix is used for **W**, then the general algorithm of Section 8.2.2 must be used.

8.4.2 The Matrix X′X

The matrix $\mathbf{X}'\mathbf{X}$ has an important role in statistics where it arises in least squares equations $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$. Properties of generalized inverses of $\mathbf{X}'\mathbf{X}$ are therefore of interest.

Theorem 8.2 *Let $(X'X)^-$ be a generalized inverse of $\mathbf{X}'\mathbf{X}$:*

(i) *$\{(X'X)^-\}'$ is also a generalized inverse of $\mathbf{X}'\mathbf{X}$.*
(ii) *$X(X'X)^-X'X = X$; that is, $(X'X)^-X'$ is a generalized inverse of X.*
(iii) *$X(X'X)^-X'$ is invariant to $(X'X)^-$.*
(iv) *$X(X'X)^-X'$ is symmetric, whether $(X'X)^-$ is symmetric or not.*

Proof.

(i) is a consequence of (8.22).
(ii) Let $H = X(X'X)^-X'X - X$. Then,

$$\begin{aligned} H'H &= [X'X\{(X'X)^-\}'X' - X'][X(X'X)^-X'X - X] \\ &= X'X\{(X'X)^-\}'X'X(X'X)^-X'X - (X'X)^-\{(X'X)^-\}'X'X \\ &\quad - X'X(X'X)^-X'X + X'X \\ &= X'X(X'X)^-X'X - X'X - X'X + X'X \\ &= \mathbf{0}. \end{aligned}$$

This implies that $H = \mathbf{0}$ and therefore (ii) is true.

(iii) Let G_1 and G_2 be two generalized inverses of $X'X$. Then $XG_1X'X = XG_2X'X$, by an application of (ii). Let $L = XG_1X' - XG_2X'$. Then,

$$\begin{aligned}
LL' &= (XG_1X' - XG_2X')(XG_1'X' - XG_2'X') \\
&= XG_1X'XG_1'X' - XG_1X'XG_2'X' - XG_2X'XG_1'X' + XG_2X'XG_2'X' \\
&= XG_2X'XG_1'X' - XG_2X'XG_2'X' - XG_2X'XG_1'X' + XG_2X'XG_2'X' \\
&= \mathbf{0}.
\end{aligned}$$

This implies that $L = \mathbf{0}$. Hence, (iii) must be true.

(iv) Define S as $S = \frac{1}{2}[(X'X)^- + \{(X'X)^-\}']$. Then,

$$X'XSX'X = \frac{1}{2}[2\,X'X] = X'X.$$

Thus, S is a symmetric generalized inverse of $X'X$. But, by (iii), $X(X'X)^-X' = XSX'$. This implies that $X(X'X)^-X$ is symmetric.

■

REFERENCES

Albert, A. (1972). *Regression and the Moore-Penrose Inverse*. Academic Press, New York.

Ben-Israel, A. and Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications*. John Wiley & Sons, New York.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, CA.

Gruber, M. H. J. (2014). *Matrix Algebra for Linear Models*. John Wiley & Sons, Hoboken, NJ.

Harville, D. A. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *Ann. Stat.*, 4, 384–395.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

Henderson, H. V. and Searle, S. R. (1980). On deriving the inverse of a sum of matrices. Paper No. BU-647-M in the Biometrics Unit Series, Cornell University, Ithaca, NY.

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26, 394–395.

Nashed, M. Z, (ed.) (1976). *Generalized Inverses and Applications*: *Proceedings of an Advanced Seminar*, University of Wisconsin, Academic Press, New York.

Penrose, R. A. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc*, 51, 406–413.

Pringle, R. M. and Rayner, A. A. (1971). *Generalized Inverse Matrices With Applications to Statistics*. Hafner Publishing Company, New York.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. John Wiley & Sons, New York.

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

EXERCISES

8.1 Show that (8.2) and (8.4) satisfy (8.1).

8.2 Show that (8.12) satisfies (8.5).

8.3 Show that (8.20) satisfies (8.19).

8.4 For $\mathbf{G}$ satisfying $\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$ show that $\mathbf{G}\mathbf{X}'$ is a generalized inverse of $\mathbf{X}$ satisfying three of the Penrose conditions.

8.5 Calculate the Moore–Penrose generalized inverses of the following matrices:

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & -1 \\ 5 & 8 & 0 & 1 \\ 1 & 2 & -2 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 2 \\ 7 & 8 & 10 & 7 \\ 2 & 1 & 1 & 6 \end{bmatrix}$$

Check your result in each case.

8.6 Find a generalized inverse of each of the following matrices:

(a) $\mathbf{PAQ}$ when $\mathbf{P}$ and $\mathbf{Q}$ are nonsingular.

(b) A^-A when A^- is a generalized inverse of $\mathbf{A}$.

(c) $k\mathbf{A}$ when k is a scalar not equal to zero.

(d) $\mathbf{ABA}'$ when $\mathbf{A}$ is orthogonal and $\mathbf{B}$ is idempotent.

(e) $\mathbf{J}$ when $\mathbf{J}$ is square with every element unity.

8.7 Prove that $\mathbf{B}^-\mathbf{A}^-$ is a generalized inverse of $\mathbf{AB}$ if and only if $\mathbf{A}^-\mathbf{ABB}^-$ is idempotent.

8.8 Prove that $r(A_r^-) = r(A)$ if A_r^- is a reflexive generalized inverse of A.

8.9 When $\mathbf{K}$ is idempotent and $\mathbf{Z}^-$ is a generalized inverse of $\mathbf{Z} = \mathbf{KAK}$, prove that $\mathbf{KZ}^-\mathbf{K}$ is also a generalized inverse of $\mathbf{Z}$.

8.10 When $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$, with $\mathbf{A}$ being symmetric, show that $\mathbf{G}^2$ is a generalized inverse of $\mathbf{A}^2$ if $\mathbf{GA}$ is symmetric.

8.11 Suppose that the matrix $\boldsymbol{P}$ is of order $m \times q$ and rank m. Show that $\boldsymbol{PP}^- = \boldsymbol{I}_m$.

8.12 If $\boldsymbol{P}_{m\times q}$ and $\boldsymbol{D}_{m\times m}$ have rank m, show that $\boldsymbol{D}^{-1} = \mathbf{P}(\mathbf{P}'\mathbf{DP})^-\mathbf{P}'$.

8.13 If $\mathbf{G}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$ define:

$$\mathbf{b} = \mathbf{G}\mathbf{X}'\mathbf{y}, \qquad s = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Show that $s = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y}$.

8.14 Suppose that $\boldsymbol{M}$ is idempotent and $\boldsymbol{V}$ is nonsingular with $\mathbf{MV} = \mathbf{VM}'$. Show that each of the following matrices is a generalized inverse of $\boldsymbol{MV}$:

(a) $\boldsymbol{M}'(\boldsymbol{MVM}')^-\boldsymbol{M}$,

(b) $\boldsymbol{V}^{-1}\boldsymbol{M}$,

(c) $\boldsymbol{M}'(\boldsymbol{MVM}')^-$.

8.15 For $\mathbf{A}$ and $\mathbf{X}$ symmetric with $\mathbf{AX} = \mathbf{0}$, and for $\mathbf{X}$ idempotent and $\mathbf{A} + \mathbf{X}$ nonsingular, prove that

(a) $(\boldsymbol{A} + \boldsymbol{X})^{-1}$ is a generalized inverse of $\boldsymbol{X}$.

(b) $\boldsymbol{A}(\boldsymbol{A} + \boldsymbol{X})^{-1}\boldsymbol{X} = \mathbf{0}$.

(c) $(\boldsymbol{A} + \boldsymbol{X})^{-1}$ is a generalized inverse of $\boldsymbol{A}$.

8.16 For $\mathbf{X}$ partitioned as $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$ with $\mathbf{X}_1$ of full column rank, prove that $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{X}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

8.17 The matrix

$$\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$$

has a number of features that are useful in statistical applications. For $\mathbf{X}$ having n rows, verify that $\mathbf{P}$ has the following properties:

(i) $\mathbf{P}$ is symmetric, idempotent and invariant to $(\mathbf{X}'\mathbf{X})^{-}$.

(ii) $\mathbf{PX} = \mathbf{0}$ and $\mathbf{X}'\mathbf{P} = \mathbf{0}$.

(iii) $r(\boldsymbol{P}) = n - r(\boldsymbol{X})$.

(iv) Columns of $\mathbf{P}$ are orthogonal to and linearly independent of columns of $\mathbf{X}$.

(v) All n-order vectors are linear combinations of the columns of $\mathbf{X}$ and $\mathbf{P}$. [Columns of $\mathbf{X}$ and $\mathbf{P}$ span Euclidean n-space and the column (row) space of $\mathbf{P}$ is the orthogonal complement of the column space of $\mathbf{X}$].

(vi) Any $\boldsymbol{B}$ for which $\boldsymbol{BX} = \mathbf{0}$ has rows that are linear combinations of the columns (rows) of $\boldsymbol{P}$; hence $\boldsymbol{B} = \boldsymbol{KP}$ for some $\boldsymbol{K}$ (the row space of $\boldsymbol{B}$ is in the column space of $\boldsymbol{P}$, the orthogonal complement of the column space of $\boldsymbol{X}$.)

(vii) $\boldsymbol{P}$ has zero row sums when row sums of $\boldsymbol{X}$ are all the same and nonzero.

(viii) $\boldsymbol{P}$ has zero row sums when a column of $\boldsymbol{X}$ is $\mathbf{1}$.

(ix) $\boldsymbol{I} - (1/n)\boldsymbol{J}_n$ is a special case of $\boldsymbol{P}$, with $\boldsymbol{X} = \mathbf{1}_n$.

(x) If $\boldsymbol{A}$ is symmetric and satisfies $\boldsymbol{AX} = \mathbf{0}$, then $\boldsymbol{A} = \boldsymbol{PLP}$ for some $\boldsymbol{L}$.

8.18 Let $\boldsymbol{M}$ be a matrix of order $n \times p$ and rank p $(n \geq p)$. Let $\boldsymbol{X}$ be defined as $\boldsymbol{X} = \boldsymbol{MT}$.

(a) Show that $\boldsymbol{X}$ and $\boldsymbol{T}$ have the same rank.

(b) Show that $\boldsymbol{M}'(\boldsymbol{MM}')^{-}\boldsymbol{M} = \boldsymbol{I}_p$.

(c) Show that $\boldsymbol{X}$ and $\boldsymbol{X}'(\boldsymbol{MM}')^{-}\boldsymbol{X}$ have the same rank.

8.19 Let $\boldsymbol{A}$ be any matrix. Show that

(a) $r(\boldsymbol{A}^{-}) \geq r(\boldsymbol{A})$.

(b) $r(\boldsymbol{A}^{-}\boldsymbol{A}) = r(\boldsymbol{AA}^{-}) = r(\boldsymbol{A})$.

(c) $\boldsymbol{A}^{-}\boldsymbol{A}$ is idempotent.

8.20 Suppose that $\boldsymbol{A}$ has p columns. Show that the general solution of $\boldsymbol{y} = \boldsymbol{Ax}$ is given by

$$\tilde{\boldsymbol{x}} = \boldsymbol{A}^{-}\boldsymbol{y} + (\boldsymbol{A}^{-}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{z},$$

where $\boldsymbol{z}$ is an arbitrary vector of p elements.

8.21 Show that the necessary and sufficient condition that $\boldsymbol{Bx}$ be unique for all $\boldsymbol{x}$ satisfying $\boldsymbol{Ax} = \boldsymbol{y}$ is that $\boldsymbol{BA}^{-}\boldsymbol{A} = \boldsymbol{B}$.

8.22 Show that $\boldsymbol{B}^{-}\boldsymbol{A}^{-}$ is a generalized inverse of $\boldsymbol{AB}$ if and only if $\boldsymbol{A}^{-}\boldsymbol{ABB}^{-}$ is idempotent.

8.23 Let $\boldsymbol{C}$ be an arbitrary matrix of order $n \times m$ and let $\boldsymbol{A}$ be a nonsingular matrix of order $n \times n$. Given that $\boldsymbol{A} + \boldsymbol{C}\boldsymbol{C}'$ is not necessarily nonsingular, show that

$$(\boldsymbol{A} + \boldsymbol{C}\boldsymbol{C}')^{-} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{C}(\mathbf{1} + \boldsymbol{C}'\boldsymbol{A}^{-1}\boldsymbol{C})^{-}\boldsymbol{C}'\boldsymbol{A}^{-1}.$$

(Hint: See Henderson and Searle, 1980, p. 15, and Harville, 1976.)

8.24 Consider the linear model,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}$ is an $n \times p$ $(n \geq p)$ matrix of rank r $(< p)$ and $\boldsymbol{\epsilon}$ is a random experimental error with mean $\mathbf{0}$ and a variance-covariance matrix $\sigma^2\boldsymbol{I}_n$. The linear function $\boldsymbol{a}'\boldsymbol{\beta}$ is said to be estimable if there exists a linear function, $\boldsymbol{c}'\boldsymbol{y}$, such that its expected value is equal to $\boldsymbol{a}'\boldsymbol{\beta}$.

(a) Show that the random vector $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ satisfies the equation

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

(b) Show that if $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}'$ belongs to the row space of $\boldsymbol{X}$.

(c) Show that if $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}'\hat{\boldsymbol{\beta}}$ is invariant to $(\boldsymbol{X}'\boldsymbol{X})^{-}$.

9

Matrix Calculus

This chapter discusses certain operations on matrices that resemble those used in single variable and multivariable calculus. These include functions of matrices, matrix differentiation, matrix inequalities, and extrema of quadratic forms. Some applications in statistics of such operations are mentioned.

9.1 MATRIX FUNCTIONS

9.1.1 Function of Matrices

Suppose that $f(x)$ is a function of a single variable x. If x is replaced with a matrix $\boldsymbol{A}$, then $f(\boldsymbol{A})$ is said to be a *matrix function*. For example, the exponential function $\exp(\boldsymbol{A})$, where $\boldsymbol{A}$ is an $n \times n$ matrix, is defined by adapting the power series for a scalar,

$$\exp(x) = \sum_{i=0}^{\infty} \frac{1}{i!} x^i,$$

to define the matrix function, $\exp(\boldsymbol{A})$, as

$$\exp(\boldsymbol{A}) = \boldsymbol{I}_n + \sum_{i=1}^{\infty} \frac{1}{i!} \boldsymbol{A}^i. \tag{9.1}$$

This representation is meaningful provided that the infinite series in (9.1) is convergent. To understand this type of convergence, the following definitions are needed:

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

Definition 9.1 *Let $\boldsymbol{A}$ be a matrix of order $m \times n$. A* norm *of $\boldsymbol{A}$, denoted by $\| \boldsymbol{A} \|$, is a scalar function with the following properties:*

(a) $\| \boldsymbol{A} \| \geq 0$, *and* $\| \boldsymbol{A} \| = 0$ *if and only if* $\boldsymbol{A} = \boldsymbol{0}$.
(b) $\| c\boldsymbol{A} \| = | c | \| \boldsymbol{A} \|$, *where c is a scalar.*
(c) $\| \boldsymbol{A} + \boldsymbol{B} \| \leq \| \boldsymbol{A} \| + \| \boldsymbol{B} \|$, *where $\boldsymbol{B}$ is any matrix of order $m \times n$.*
(d) $\| \boldsymbol{AC} \| \leq \| \boldsymbol{A} \| \| \boldsymbol{C} \|$, *where $\boldsymbol{C}$ is any matrix for which the product $\boldsymbol{AC}$ is defined.*

For example, the *Euclidean norm*, $\| \boldsymbol{A} \|_2$, is defined as

$$\| \boldsymbol{A} \|_2 = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}}, \tag{9.2}$$

where $\boldsymbol{A} = (a_{ij})$ is a matrix of order $m \times n$. Another example of a matrix norm is the *spectral norm*,

$$\| \boldsymbol{A} \|_s = \left[e_{max}(\boldsymbol{A}'\boldsymbol{A}) \right]^{\frac{1}{2}}, \tag{9.3}$$

where $e_{max}(\boldsymbol{A}'\boldsymbol{A})$ is the largest eigenvalue of $\boldsymbol{A}'\boldsymbol{A}$.

Definition 9.2 *Let $\{\boldsymbol{A}_k\}_{k=1}^{\infty}$ be an infinite sequence of matrices of order $m \times n$. The infinite series $\sum_{k=1}^{\infty} \boldsymbol{A}_k$ converges to the $m \times n$ matrix $\boldsymbol{S} = (s_{ij})$ if the series $\sum_{k=1}^{\infty} a_{ijk}$ converges for all $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, where a_{ijk} is the* (i, j)th *element of $\boldsymbol{A}_k$, and*

$$\sum_{k=1}^{\infty} a_{ijk} = s_{ij}, \quad i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n. \tag{9.4}$$

The series $\sum_{k=1}^{\infty} \boldsymbol{A}_k$ is divergent if at least one of the series in (9.4) is divergent.

Theorem 9.1 *Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$ and let λ_i be its* ith *eigenvalue ($i = 1, 2, \ldots, n$). If $|\lambda_i| < 1$ for $i = 1, 2, \ldots, n$, then $\sum_{k=0}^{\infty} \boldsymbol{A}^k$ converges to $(\boldsymbol{I} - \boldsymbol{A})^{-1}$.*

Proof. Using the spectral decomposition theorem (Theorem 6.2) we can write $\boldsymbol{A} = \boldsymbol{P\Lambda P}'$, where $\boldsymbol{P}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{A}$ (Note that all these eigenvalues are real by Lemma 6.1). We then have

$$\boldsymbol{A}^k = \boldsymbol{P\Lambda}^k\boldsymbol{P}', \quad k = 0, 1, \ldots$$

It can be seen that $\boldsymbol{A}^k \to \boldsymbol{0}$ as $k \to \infty$ since $\boldsymbol{\Lambda}^k \to \boldsymbol{0}$ by the fact that $|\lambda_i| < 1$ for all i. Furthermore, the matrix $\boldsymbol{I} - \boldsymbol{A}$ is nonsingular since $\boldsymbol{I} - \boldsymbol{A} = \boldsymbol{P}(\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{P}'$ and the diagonal elements $\boldsymbol{I} - \boldsymbol{\Lambda}$ are positive.

Now, on the basis of Section 4.12, the following identity holds for any non-negative integer k:

$$\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^k = (\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} - \boldsymbol{A}^{k+1}).$$

Letting k go to ∞, we get

$$\sum_{k=0}^{\infty} \boldsymbol{A}^k = (\boldsymbol{I} - \boldsymbol{A})^{-1}. \tag{9.5}$$

■

Theorem 9.2 *Let* $\boldsymbol{A}$ *be a symmetric matrix of order* $n \times n$ *and* λ *be any eigenvalue of* $\boldsymbol{A}$*. If* $\| \boldsymbol{A} \|$ *is any matrix norm of* $\boldsymbol{A}$*, then* $|\lambda| \leq \| \boldsymbol{A} \|$.

Proof. Let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{A}$ corresponding to λ. Then, $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$. Furthermore,

$$\| \lambda\boldsymbol{v} \| = |\lambda| \; \| \boldsymbol{v} \| = \| \boldsymbol{A}\boldsymbol{v} \| \leq \| \boldsymbol{A} \| \| \boldsymbol{v} \| .$$

This implies $|\lambda| \leq \| \boldsymbol{A} \|$ since $\boldsymbol{v} \neq \boldsymbol{0}$. ■

Corollary 9.1 *Let* $\boldsymbol{A}$ *be a symmetric matrix of order* $n \times n$ *such that* $\| \boldsymbol{A} \| < 1$*, where* $\| \boldsymbol{A} \|$ *is any matrix norm of* $\boldsymbol{A}$*. Then* $\sum_{k=0}^{\infty} \boldsymbol{A}^k$ *converges to* $(\boldsymbol{1} - \boldsymbol{A})^{-1}$.

Proof. This follows directly from Theorem 9.1 since $|\lambda_i| \leq \| \boldsymbol{A} \| < 1$ for $i = 1, 2, \ldots, n$ by Theorem 9.2. ■

For other matrix functions such as $\exp(\boldsymbol{A})$, $\sin(\boldsymbol{A})$, and $\cos(\boldsymbol{A})$ we have the following series representations:

$$\begin{aligned}
\exp(\boldsymbol{A}) &= \sum_{k=0}^{\infty} \frac{1}{k!}\boldsymbol{A}^k \\
\sin(\boldsymbol{A}) &= \sum_{k=0}^{\infty} (-1)^k \frac{1}{(2k+1)!}\boldsymbol{A}^{2k+1} \\
\cos(\boldsymbol{A}) &= \sum_{k=0}^{\infty} (-1)^k \frac{1}{(2k)!}\boldsymbol{A}^{2k}.
\end{aligned}$$

Another series representation of a well-known matrix function is given by

$$\log(\boldsymbol{I} - \boldsymbol{A}) = \sum_{k=1}^{\infty} \frac{1}{k}\boldsymbol{A}^k,$$

where $\boldsymbol{A}$ is a symmetric matrix whose eigenvalues fall inside the open interval $(-1, 1)$.

9.1.2 Matrices of Functions

Matrices can have elements that are functions of one variable x, or of several variables $x_1, x_2, \ldots, x_r$, which can be represented as $\boldsymbol{x}$. Using $\mathbf{F}$ to represent matrices of such elements, their functional relationship to x or $\boldsymbol{x}$ can be represented by $\mathbf{F}(x)$ and $\mathbf{F}(\boldsymbol{x})$, respectively. Simple examples are

$$\mathbf{F}(x) = \begin{bmatrix} x^3 & e^x \\ 2^x & 3+4x^2 \end{bmatrix}$$

and, for $\quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{F}(\boldsymbol{x}) = \begin{bmatrix} x_1^2 & x_1x_2 & x_2^2 \\ x_1x_2 & x_2^3 & x_1+x_2 \end{bmatrix}.$

This notation is used in the next section.

9.2 ITERATIVE SOLUTION OF NONLINEAR EQUATIONS

Matrices can be used in formulating iterative solutions of nonlinear equations, as is illustrated by the following elementary development of one particular method.

Suppose equations to be solved are

$$f_i(x_1, x_2, \ldots, x_n) = 0 \quad \text{for} \quad i = 1, 2, \ldots, n,$$

where $f_i(x_1, x_2, \ldots, x_n)$ is a nonlinear function of n unknowns $x_1,\ x_2, \ldots, x_n$. Denoting the $x's$ by $\boldsymbol{x}$ and the vector of functions $f_i(\boldsymbol{x})$ by $\boldsymbol{f}(\boldsymbol{x})$, the equations can be written succinctly as

$$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}. \tag{9.6}$$

Define the $n \times n$ matrix

$$\boldsymbol{G}(\boldsymbol{x}) = \{g_{ij}(\boldsymbol{x})\} = \left\{ \frac{\partial}{\partial x_j} f_i(\boldsymbol{x}) \right\} \quad \text{for} \quad i,j = 2, 2, \ldots, n.$$

We suppose the functions $f_i(x)$ are sufficiently complicated that (9.6) cannot be solved explicitly for $\boldsymbol{x}$. Let $\boldsymbol{x}_0$ be some first guess at a solution. Then a succession of improvements on that guess (that is, successive iterative solutions) with $\boldsymbol{x}_r$ being the rth improvement, can be derived using the equation

$$\boldsymbol{f}(\boldsymbol{x}_{r+1}) = \boldsymbol{f}(\boldsymbol{x}_r) + \boldsymbol{G}(\boldsymbol{x}_r)\boldsymbol{\Delta}_r, \tag{9.7}$$

where

$$\boldsymbol{\Delta}_r = \boldsymbol{x}_{r+1} - \boldsymbol{x}_r. \tag{9.8}$$

To a first approximation, (9.7) is a multivariable extension of Taylor's expansion.

Starting with $\boldsymbol{x}_0$, that is, with $r = 0$, we use (9.7) to derive $\boldsymbol{\Delta}_r$ and thence obtain $\boldsymbol{x}_{r+1}$ from (9.8) as $\boldsymbol{x}_{r+1} = \boldsymbol{x}_r + \boldsymbol{\Delta}_r$. This is done by noting that, were $\boldsymbol{x}_{r+1}$ a solution to (9.6), then (9.7) would be $\mathbf{0} = \boldsymbol{f}(\boldsymbol{x}_r) + \boldsymbol{G}(\boldsymbol{x}_r)\boldsymbol{\Delta}_r$ so giving

$$\boldsymbol{\Delta}_r = -\left[\boldsymbol{G}(\boldsymbol{x}_r\right]^{-1}\boldsymbol{f}(\boldsymbol{x}_r), \tag{9.9}$$

providing $\boldsymbol{G}(\boldsymbol{x}_r)$ is nonsingular. With this,

$$\boldsymbol{x}_{r+1} = \boldsymbol{x}_r + \boldsymbol{\Delta}_r. \tag{9.10}$$

Equation (9.9) is first used with $\boldsymbol{x}_r$ being given the value $\boldsymbol{x}_0$, and then (9.10) gives $\boldsymbol{x}_1$; using this for $\boldsymbol{x}_r$ in (9.9) and then (9.10) again gives $\boldsymbol{x}_2$; and so on until, for some integer r, a value for $\boldsymbol{\Delta}_r$ is obtained that is sufficiently close to $\mathbf{0}$ for us to be satisfied that $\boldsymbol{x}_r$ is, to that same degree of approximation, a solution of (9.6). Consideration must of course be given to matters such as convergence to a solution, the rate of convergence, multiple solutions, arithmetic stability, and so on. These are problems for the numerical analyst and are beyond the scope of this book. Development of (9.9) and (9.10) are given solely as illustration of the uses of matrix notation.

9.3 VECTORS OF DIFFERENTIAL OPERATORS

This section and the next make simple use of, and should be omitted by readers not familiar with, the differential calculus.

9.3.1 Scalars

Sometimes a function of several variables has to be differentiated with respect to each of the variables concerned. In such cases, it is convenient to write the derivatives as a column vector. A common use of this involves only linear and quadratic functions of variables, so we confine ourselves to these. For example, if

$$\lambda = 3x_1 + 4x_2 + 9x_3$$

where λ, x_1, x_2, and x_3 are scalars, we write the derivatives of λ with respect to x_1, x_2, and x_3 as

$$\frac{\partial \lambda}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial \lambda}{\partial x_1} \\ \dfrac{\partial \lambda}{\partial x_2} \\ \dfrac{\partial \lambda}{\partial x_3} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \dfrac{\partial}{\partial x_2} \\ \dfrac{\partial}{\partial x_3} \end{bmatrix} \lambda = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}.$$

The second vector shows how the symbol $\dfrac{\partial}{\partial \boldsymbol{x}}$ can represent a whole vector of differential operators. Noting that

$$\lambda = 3x_1 + 4x_2 + 9x_3 = [3 \quad 4 \quad 9]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{a}'\mathbf{x},$$

so defining vectors $\boldsymbol{a}$ and $\boldsymbol{x}$, we therefore see that with $\dfrac{\partial}{\partial \boldsymbol{x}}$ representing a column of operators the differential of the scalar $\boldsymbol{a}'\boldsymbol{x} = \boldsymbol{x}'\mathbf{a}$ is

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{a}) = \boldsymbol{a}. \tag{9.11}$$

9.3.2 Vectors[1]

Suppose the principle of (9.11) is applied to each element of $\boldsymbol{y}' = \boldsymbol{x}'\boldsymbol{A}$ expressed as

$$\boldsymbol{y}' = [y_1 \quad y_2 \quad \cdots \quad y_n] = [\boldsymbol{x}'\boldsymbol{a}_1 \quad \boldsymbol{x}'\boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{x}'\boldsymbol{a}_n] \tag{9.12}$$

where the ith element of $\boldsymbol{y}$ is $y_i = \boldsymbol{x}'\boldsymbol{a}_i$ for $\boldsymbol{a}_\imath$ being the ith column of $\boldsymbol{A}$, with $i = 1, \ldots, n$. Applying (9.11) to each element of (9.12) gives

$$\begin{aligned}\frac{\partial \boldsymbol{y}'}{\partial \boldsymbol{x}} &= \begin{bmatrix} \dfrac{\partial y_1}{\partial \boldsymbol{x}} & \dfrac{\partial y_2}{\partial \boldsymbol{x}} & \cdots & \dfrac{\partial y_n}{\partial \boldsymbol{x}} \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\partial \boldsymbol{x}'\boldsymbol{a}_1}{\partial \boldsymbol{x}} & \dfrac{\partial \boldsymbol{x}'\boldsymbol{a}_2}{\partial \boldsymbol{x}} & \cdots & \dfrac{\partial \boldsymbol{x}'\boldsymbol{a}_n}{\partial \boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix} = \boldsymbol{A}.\end{aligned}$$

Thus

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}) = \boldsymbol{A}. \tag{9.13}$$

In differentiating the row vector $\boldsymbol{y}' = \boldsymbol{x}'\boldsymbol{A}$ the columns $\partial y_\imath/\partial \boldsymbol{x}$ are, in deriving (9.13), set alongside one another. We do the same in differentiating the column vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$; otherwise, the logical procedure of putting the columns $\partial y_\imath/\partial \boldsymbol{x}$ one under the other would yield a vector having the same order as $\boldsymbol{x} \otimes \boldsymbol{y}$. The convention is therefore adopted that by $\partial(\boldsymbol{A}\boldsymbol{x})/\partial \boldsymbol{x}$ we mean $\partial(\boldsymbol{A}\boldsymbol{x})'/\partial \boldsymbol{x}$ and so, by (9.13)

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}) \equiv \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x})' = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}') = \boldsymbol{A}'. \tag{9.14}$$

For example, for $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A}$ of order 3×3,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_3}{\partial x_1} \\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} & \dfrac{\partial y_3}{\partial x_2} \\ \dfrac{\partial y_1}{\partial x_3} & \dfrac{\partial y_2}{\partial x_3} & \dfrac{\partial y_3}{\partial x_3} \end{bmatrix} = \boldsymbol{A}'. \tag{9.15}$$

Example 9.1 *Applying (9.15) to*

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 2 & 6 & -1 \\ 3 & -2 & 4 \\ 3 & 4 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 - x_3 \\ 3x_1 - 2x_2 + 4x_3 \\ 3x_1 + 4x_2 + 7x_3 \end{bmatrix}$$

[1] Thanks go to G. P. H. Styan for suggesting improved notation for this section.

gives $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{Ax})$

$$= \left[\frac{\partial}{\partial \mathbf{x}}(2x_1 + 6x_2 - x_3) \quad \frac{\partial}{\partial \mathbf{x}}(3x_1 - 2x_2 + 4x_3) \quad \frac{\partial}{\partial \mathbf{x}}(3x_1 + 4x_2 + 7x_3)\right]$$

$$= \begin{bmatrix} 2 & 3 & 3 \\ 6 & -2 & 4 \\ -1 & 4 & 7 \end{bmatrix} = \boldsymbol{A}'.$$

9.3.3 Quadratic Forms

Adapting the principle of differentiation of the dot product, $\boldsymbol{f}'\boldsymbol{g}$, with respect to $\boldsymbol{x}$, namely,

$$\frac{\partial(\boldsymbol{f}'\boldsymbol{g})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{f}'}{\partial \boldsymbol{x}}\boldsymbol{g} + \frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{x}}\boldsymbol{f},$$

where $\boldsymbol{f} = \boldsymbol{x}$ and $\boldsymbol{g} = \boldsymbol{Ax}$, the differentiation of $\boldsymbol{x}'\boldsymbol{Ax}$ with respect to $\boldsymbol{x}$ is

$$\begin{aligned} \frac{\partial(\boldsymbol{x}'\boldsymbol{Ax})}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{x}'}{\partial \boldsymbol{x}}\boldsymbol{Ax} + \frac{\partial(\boldsymbol{Ax})'}{\partial \boldsymbol{x}}\boldsymbol{x} \\ &= \boldsymbol{Ax} + \boldsymbol{A}'\boldsymbol{x}, \quad \text{from (9.13) and (9.14).} \end{aligned} \tag{9.16}$$

In particular, if $\boldsymbol{A}$ is symmetric, then

$$\frac{\partial(\boldsymbol{x}'\boldsymbol{Ax})}{\partial \boldsymbol{x}} = 2\boldsymbol{Ax}. \tag{9.17}$$

Example 9.2 *For*

$$\begin{aligned} q = \boldsymbol{x}'\boldsymbol{Ax} &= [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 1 & 3 & 5 \\ 3 & 4 & 7 \\ 5 & 7 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= x_1^2 + 6x_1x_2 + 10x_1x_3 + 4x_2^2 + 14x_2x_3 + 9x_3^2, \end{aligned}$$

$$\frac{\partial q}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial q}{\partial x_1} \\ \dfrac{\partial q}{\partial x_2} \\ \dfrac{\partial q}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 + 10x_3 \\ 6x_1 + 8x_2 + 14x_3 \\ 10x_1 + 14x_2 + 18x_3 \end{bmatrix} = 2\boldsymbol{Ax}.$$

Example 9.3 *The method of least squares in statistics involves minimizing the sum of squares of the elements of the vector* $\boldsymbol{e}$ *in the equation*

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{e}. \tag{9.18}$$

If $\mathbf{e} = \{e_i\}$ *for* $i = 1, 2, \ldots, n$*, the sum of squares is*

$$\begin{aligned} S &= \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}, \end{aligned}$$

and because $\mathbf{y}'\mathbf{X}\mathbf{b}$ *is a scalar it equals its transpose,* $\mathbf{b}'\mathbf{X}\mathbf{y}$*, and hence*

$$S = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}.$$

Minimizing this with respect to elements of $\mathbf{b}$ *involves equating to zero the expression* $\dfrac{\partial S}{\partial \mathbf{b}}$*, where* $\dfrac{\partial}{\partial \mathbf{b}}$ *is a vector of differential operators of the nature just discussed. Using results (9.13) and (9.17) gives*

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b},$$

and equating this to a null vector leads to the equations

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \tag{9.19}$$

with solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}. \tag{9.20}$$

Example 9.4 *The selection index method of genetics as proposed by Hazel (1943) can be formulated as follows: Let* $\mathbf{x}$ *be a vector of observations (phenotypes) on certain traits in an animal and* $\mathbf{b}$ *a vector of coefficients by which these observations get combined into an index* $I = \mathbf{b}'\mathbf{x}$*. Corresponding to the traits represented in* $\mathbf{x}$*, let* $\mathbf{g}$ *be a vector of (unknown) genetic values and let* $\mathbf{a}$ *be a vector of the relative economic values of these traits. Then the coefficients represented by* $\mathbf{b}$ *are determined by maximizing, for values of* $\mathbf{b}$*, the correlation between* $I = \mathbf{b}'\mathbf{x}$ *and the combined genetic value* $H = \mathbf{a}'\mathbf{g}$*. This correlation is*

$$r_{IH} = \frac{\text{cov}(IH)}{\sigma_I \sigma_H} = \frac{\mathbf{b}'\mathbf{G}\mathbf{a}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}\sqrt{\mathbf{a}'\mathbf{G}\mathbf{a}}},$$

where $\mathbf{P}$ *is the matrix of phenotypic variances and covariances, namely the variance-covariance matrix appropriate to the vector of phenotypic values* $\mathbf{x}$*; and* $\mathbf{G}$ *is the matrix of genetic variances and covariances, the variance-covariance matrix corresponding to the genetic values* $\mathbf{g}$*. Maximizing this with respect to the elements in* $\mathbf{b}$ *is achieved by maximizing* $\mathbf{b}'\mathbf{G}\mathbf{a}/\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}$*. Equating to zero the derivative of this with respect to* $\mathbf{b}$*, using results (9.13) and (9.17) to carry out the differentiation,* $\mathbf{P}$ *and* $\mathbf{G}$ *being symmetric, gives*

$$\left(\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}\right)\mathbf{G}\mathbf{a} = \frac{1}{2}(\mathbf{b}'\mathbf{P}\mathbf{b})^{-\frac{1}{2}}2\mathbf{P}\mathbf{b}(\mathbf{b}'\mathbf{G}\mathbf{a}),$$

that is,

$$(\mathbf{b}'\mathbf{G}\mathbf{a})\mathbf{P}\mathbf{b} = (\mathbf{b}'\mathbf{P}\mathbf{b})\mathbf{G}\mathbf{a}.$$

Apart from a scalar, all solutions to this equation are proportional to solutions to $\mathbf{Pb} = \mathbf{Ga}$. The required value of $\mathbf{b}$ is therefore taken as the solution of the latter equation:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Ga}. \tag{9.21}$$

The corresponding (maximized) value of r_{IH} is then $r_{IH} = \sqrt{\dfrac{\mathbf{b}'\mathbf{Ga}}{\mathbf{a}'\mathbf{Ga}}}$.

Estimates of $\mathbf{b}$ and r_{IH} are obtained by using a predetermined value of $\mathbf{a}$ and estimates of the phenotypic and genetic variances and covariances involved in $\mathbf{P}$ and $\mathbf{G}$.

9.4 VEC AND VECH OPERATORS

9.4.1 Definitions

A matrix operation dating back nearly a century, but in which there is currently great renewed interest [see Henderson and Searle(1979)], is that of stacking the columns of a matrix one under the other to form a single column. Over the years it has had a variety of names, the most recent being "vec." Thus for

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad \text{vec}\,\mathbf{X} = \begin{bmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{bmatrix}.$$

The notations vec $\mathbf{X}$ and vec($\mathbf{X}$) are used interchangeably, the latter only when parentheses are deemed necessary.

An extension of vec $\mathbf{X}$ is vech $\mathbf{X}$, defined in the same way that vec $\mathbf{X}$ is, except that for each column of $\mathbf{X}$ only that part of it which is on or below the diagonal of $\mathbf{X}$ is put into vech $\mathbf{X}$ (vector-half of $\mathbf{X}$). In this way, for $\mathbf{X}$ symmetric, vech $\mathbf{X}$ contains only the distinct elements of $\mathbf{X}$; that is, for

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 6 \\ 7 & 3 & 8 \\ 6 & 8 & 2 \end{bmatrix} = \mathbf{X}', \qquad \text{vech}\,\mathbf{X} = \begin{bmatrix} 1 \\ 7 \\ 6 \\ 3 \\ 8 \\ 2 \end{bmatrix}.$$

Henderson and Searle(1979) give history, properties, and many applications of these operators. Only a few highlights are summarized here.

9.4.2 Properties of Vec

The following three theorems give useful properties of the vec operator. Proofs of the first two depend on the elementary vector $\mathbf{e}_i$, the ith column of an identity matrix, discussed in Section 5.1.5.

Theorem 9.3 *vec*$(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})$*vec* $\mathbf{B}$.

Proof. The jth column of $\mathbf{ABC}$, and hence the jth subvector of vec($\mathbf{ABC}$) is, for $\mathbf{C}$ having r rows,

$$\mathbf{ABCe}_j = \mathbf{AB}\sum_i \mathbf{e}_i\mathbf{e}_i'\mathbf{Ce}_j = \sum_i c_{ij}\mathbf{ABe}_i = \begin{bmatrix} c_{1j}\mathbf{A} & c_{2j}\mathbf{A} & \cdots & c_{rj}\mathbf{A}\end{bmatrix}\begin{bmatrix}\mathbf{Be}_1\\ \mathbf{Be}_2\\ \vdots\\ \mathbf{Be}_r\end{bmatrix},$$

which is the jth subvector of $\mathbf{C}' \otimes \mathbf{A}$)vec $\mathbf{B}$. ∎

Theorem 9.4 *tr*$(\mathbf{AB}) = ($*vec* $\mathbf{A}')'$ *vec* $\mathbf{B}$.

Proof. $tr(\mathbf{AB}) = \Sigma_i \mathbf{e}_i'\mathbf{ABe}_i = [\mathbf{e}_1'\mathbf{A} \quad \cdots \quad \mathbf{e}_r'\mathbf{A}]\begin{bmatrix}\mathbf{Be}_1\\ \vdots\\ \mathbf{Be}_r\end{bmatrix} = (\text{vec}\,\mathbf{A}')'\text{vec}\,\mathbf{B}.$ ∎

Theorem 9.5 *tr*$(\mathbf{AZ}'\mathbf{BZC}) = ($*vec* $\mathbf{Z})'(\mathbf{CA} \otimes \mathbf{B}')$*vec* $\mathbf{Z}$.

Proof.

$$\begin{aligned} tr(\mathbf{AZ}'\mathbf{BZC}) &= tr(\mathbf{Z}'\mathbf{BZCA}) \\ &= (\text{vec}\,\mathbf{Z})'\text{vec}(\mathbf{BZCA}), \quad \text{from Theorem 9.4} \\ &= (\text{vec}\,\mathbf{Z})'(\mathbf{A}'\mathbf{C}' \otimes \mathbf{B})\text{vec}\,\mathbf{Z}, \quad \text{from Theorem 9.3,} \end{aligned}$$

and, being a scalar, this equals its own transpose. ∎

9.4.3 Vec-Permutation Matrices

The vectors vec $\mathbf{A}$ and vec $\mathbf{A}'$ contain the same elements, but in different sequences. We define $\mathbf{I}_{m,n}$ as that permutation matrix (see Section 4.17) which rearranges vec $\mathbf{A}'$ to be vec $\mathbf{A}$:

$$\text{vec}(\mathbf{A}_{m\times n}) = \mathbf{I}_{m,n}\text{vec}(\mathbf{A}'). \tag{9.22}$$

$\mathbf{I}_{m,n}$ is called a *vec-permutation* matrix.

The nature of the relationship between vec $\mathbf{A}$ and vec $\mathbf{A}'$ yields a variety of characteristics for the vec-permutation matrix, not the least of which is the part played in connecting the direct products (Section 4.11) $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{B} \otimes \mathbf{A}$:

$$\mathbf{B}_{p\times q} \otimes \mathbf{A}_{m\times n} = \mathbf{I}_{m,p}(\mathbf{A}_{m\times n} \otimes \mathbf{B}_{p\times q})\mathbf{I}_{q,n}. \tag{9.23}$$

This and more, including the many names, notations and definitions given to $\mathbf{I}_{m,n}$ during the last 50 years, are presented at length in Henderson and Searle (1981b).

9.4.4 Relationships Between Vec and Vech

When $\mathbf{X}$ is symmetric, the elements of vec $\mathbf{X}$ are those of vech $\mathbf{X}$ with some repetitions. Therefore vec $\mathbf{X}$ and vech $\mathbf{X}$ for symmetric $\mathbf{X}$ are linear transformations of one another:

$$\text{for } \mathbf{X} = \mathbf{X}', \quad \text{vech}\,\mathbf{X} = \mathbf{H}\,\text{vec}\,\mathbf{X}, \quad \text{and} \quad \text{vec}\,\mathbf{X} = \mathbf{G}\,\text{vech}\,\mathbf{X}, \tag{9.24}$$

so defining $\mathbf{H}$ and $\mathbf{G}$. They have the following properties, for $\mathbf{X}$ being $n \times n$:

$\mathbf{G}$ is unique, of full column rank $\frac{1}{2}n(n+1)$, with $\mathbf{I}_{n,n}\mathbf{G} = \mathbf{G}$.
$\mathbf{H}$ is not unique, but does have full row rank $\frac{1}{2}n(n+1)$.
$\mathbf{H}$ is a left inverse of $\mathbf{G}$, that is, $\mathbf{HG} = \mathbf{I}$.

One form of $\mathbf{H}$ is $\mathbf{H} = \mathbf{G}^+ = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$, the Moore–Penrose inverse of $\mathbf{G}$, for which $\mathbf{HI}_{n,n} = \mathbf{H}$ and $\mathbf{GH} = \frac{1}{2}(\mathbf{I}_{n^2} + \mathbf{I}_{n,n})$.

For any $\mathbf{H}$, $|\mathbf{H}(X \otimes X)\mathbf{G}| = |X|^{n+1}$.

These and more are discussed and established by Henderson and Searle (1981a), who also give specific examples of $\mathbf{H}$ and $\mathbf{G}$.

9.5 OTHER CALCULUS RESULTS

The formulations in Section 9.3 use matrix notation to represent the differentiating of functions with respect to an array of variables. This principle can also be applied to a variety of functions of matrices, of which we give but a small selection here. The reader interested in further examples is referred to Deemer and Olkin (1951), Dwyer (1967), Neudecker (1969), Henderson and Searle (1979), Nel (1980), Rogers (1980), Graybill (1983, Chapter 10), and Harville (1997, Chapter 15), and to the many references they contain.

9.5.1 Differentiating Inverses

Suppose that the elements of $\mathbf{A}_{r\times c}$ are functions of a scalar x. The differential of $\mathbf{A}$ with respect to x is defined as the matrix of each element of $\mathbf{A}$ differentiated with respect to x:

$$\frac{\partial \mathbf{A}}{\partial x} = \left\{ \frac{\partial a_{ij}}{\partial x} \right\} \quad \text{for} \quad i = 1, \ldots, r \quad \text{and} \quad j = 1, \ldots, c. \tag{9.25}$$

Applying this to the product $\boldsymbol{AA}^{-1} = \boldsymbol{I}$ for nonsingular $\boldsymbol{A}$ gives

$$\frac{\partial \boldsymbol{A}^{-1}}{\partial x} = -\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1}. \tag{9.26}$$

9.5.2 Differentiating Traces

A variation on (9.17), for differentiating a scalar θ that is a function of elements of $\boldsymbol{X}_{p\times q}$, is to define

$$\frac{\partial \theta}{\partial \boldsymbol{X}} = \left\{\frac{\partial \theta}{\partial x_{ij}}\right\} \quad \text{for } i = 1, \dots, p \text{ and } j = 1, \dots, q. \tag{9.27}$$

Trace functions involving $\boldsymbol{X}$ are special cases of this. For $\boldsymbol{X}$ having functionally independent elements

$$\frac{\partial}{\partial x_{ij}} tr(\boldsymbol{XA}) = \frac{\partial}{\partial x_{ij}} \sum_r \sum_s x_{rs} a_{sr} = a_{ji}$$

and the matrix of all such results is

$$\frac{\partial}{\partial \boldsymbol{X}} tr(\boldsymbol{XA}) = \boldsymbol{A}'. \tag{9.28}$$

But (9.28) does not hold when elements of $\boldsymbol{X}$ are functionally related as, for example, when $\boldsymbol{X}$ is symmetric. Then, as is shown in Henderson and Searle [1981a, equation (58)],

$$\frac{\partial}{\partial \boldsymbol{X}} tr(\boldsymbol{XA}) = \boldsymbol{A} + \boldsymbol{A}' - \text{diag}(\boldsymbol{A}), \quad \text{for symmetric } \boldsymbol{X}, \tag{9.29}$$

where $diag(\boldsymbol{A})$ is the diagonal matrix of diagonal elements of $\boldsymbol{A}$, a variation on the notation for diagonal matrices given in Chapter 2.

9.5.3 Derivative of a Matrix with Respect to Another Matrix

Let $\boldsymbol{Y} = (y_{ij})$ be matrix of order $m \times n$ whose elements are functions of the elements of a matrix $\boldsymbol{X} = (x_{ij})$, which is of order $p \times q$. Then, the derivative of $\boldsymbol{Y}$ with respect to $\boldsymbol{X}$, denoted by $\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}}$, is given by the partitioned matrix

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial \boldsymbol{Y}}{\partial x_{11}} & \frac{\partial \boldsymbol{Y}}{\partial x_{12}} & \cdots & \frac{\partial \boldsymbol{Y}}{\partial x_{1q}} \\ \frac{\partial \boldsymbol{Y}}{\partial x_{21}} & \frac{\partial \boldsymbol{Y}}{\partial x_{22}} & \cdots & \frac{\partial \boldsymbol{Y}}{\partial x_{2q}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \boldsymbol{Y}}{\partial x_{p1}} & \frac{\partial \boldsymbol{Y}}{\partial x_{p2}} & \cdots & \frac{\partial \boldsymbol{Y}}{\partial x_{pq}} \end{bmatrix},$$

where $\dfrac{\partial \boldsymbol{Y}}{\partial x_{ij}}$ is the derivative of $\boldsymbol{Y}$ with respect to x_{ij}. Since $\dfrac{\partial \boldsymbol{Y}}{\partial x_{ij}}$ is of order $m \times n$, the matrix $\dfrac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}}$ is of order $mp \times nq$.

For example, if

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix},$$

and

$$\boldsymbol{Y} = \begin{bmatrix} x_{11} & x_{21}^3 \\ \cos(2\,x_{11} + 3\,x_{22}) & e^{x_{11}} \end{bmatrix},$$

then,

$$\frac{\partial \boldsymbol{Y}}{\partial x_{11}} = \begin{bmatrix} 1 & 0 \\ -2\sin(2\,x_{11} + 3\,x_{22}) & e^{x_{11}} \end{bmatrix}$$

$$\frac{\partial \boldsymbol{Y}}{\partial x_{12}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\frac{\partial \boldsymbol{Y}}{\partial x_{21}} = \begin{bmatrix} 0 & 3\,x_{21}^2 \\ 0 & 0 \end{bmatrix}$$

$$\frac{\partial \boldsymbol{Y}}{\partial x_{22}} = \begin{bmatrix} 0 & 0 \\ -3\sin(2\,x_{11} + 3\,x_{22}) & 0 \end{bmatrix}.$$

Hence,

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2\sin(2\,x_{11} + 3\,x_{22}) & e^{x_{11}} & 0 & 0 \\ 0 & 3\,x_{21}^2 & 0 & 0 \\ 0 & 0 & -3\sin(2\,x_{11} + 3\,x_{22}) & 0 \end{bmatrix}.$$

9.5.4 Differentiating Determinants

For $\boldsymbol{X}$, which is of order $p \times q$, having all elements functionally independent, it is clear that

$$\frac{\partial \boldsymbol{X}}{\partial x_{ij}} = \mathbf{E}_{ij}, \tag{9.30}$$

a null matrix except for unity as the (i, j)th element, the matrix $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j'$ being as defined in Section 5.1.5. Hence, if $\mathbf{Q}$ is an $m \times p$ constant matrix, then

$$\begin{aligned}\frac{\partial(\mathbf{QX})}{\partial x_{ij}} &= \mathbf{Q}\frac{\partial \mathbf{X}}{\partial x_{ij}} \\ &= \mathbf{QE}_{ij}, \ \ i = 1, 2, \ldots, p; \, j = 1, 2, \ldots, q. \end{aligned} \tag{9.31}$$

Thus if $\dfrac{\partial \mathbf{QX}}{\partial \mathbf{X}} = \mathbf{0}$, then $\mathbf{Q}$ must be equal to a zero matrix (see Exercise 19).

Similar results apply for symmetric $\mathbf{X}$, for which

$$\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{E}_{ij} + \mathbf{E}_{ji} - \delta_{ij}\mathbf{E}_{ij} \tag{9.32}$$

for $\delta_{ij} = 0$ when $i \neq j$ and $\delta_{ij} = 1$ when $i = j$.

Recalling from Chapter 3 that $|\mathbf{X}| = \Sigma_j x_{ij}|\mathbf{X}_{ij}|$ for each i, where $\mathbf{X} = \{x_{ij}\}$ and $|\mathbf{X}_{ij}|$ is the cofactor of x_{ij}, we have, in contrast to (9.30),

$$\frac{\partial|\mathbf{X}|}{\partial x_{ij}} = |\mathbf{X}_{ij}| \quad \text{for } \mathbf{X} \text{ having functionally independent elements.} \tag{9.33}$$

But, similar to (9.32),

$$\begin{aligned}\frac{\partial|\mathbf{X}|}{\partial x_{ij}} &= |\mathbf{X}_{ij}| + |\mathbf{X}_{ji}| - \delta_{ij}|\mathbf{X}_{ij}| \quad \text{for symmetric } \mathbf{X}, \\ &= (2 - \delta_{ij})|\mathbf{X}_{ij}| \quad \text{for symmetric } \mathbf{X}. \end{aligned} \tag{9.34}$$

Therefore, on denoting the adjoint (or adjugate) matrix of $\mathbf{X}$ by adj $\mathbf{X}$ (see Section 4.12), and on arraying results (9.33) and (9.34) as matrices, we have, for $\mathbf{X}^{-1} = \text{adj}(\mathbf{X})/|\mathbf{X}|$,

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = (\text{adj}\,\mathbf{X})' = |\mathbf{X}|(\mathbf{X}^{-1})', \text{ for } \mathbf{X} \text{ having functionally unrelated elements} \tag{9.35}$$

and

$$\begin{aligned}\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} &= 2(\text{adj}\,\mathbf{X}) - \text{diag}(\text{adj}\,\mathbf{X}) \\ &= |\mathbf{X}|[2\mathbf{X}^{-1} - \text{diag}(\mathbf{X}^{-1})], \text{ for symmetric } \mathbf{X}. \end{aligned} \tag{9.36}$$

Furthermore, for nonsingular $\mathbf{X}$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}} \log|\mathbf{X}| &= \frac{1}{|\mathbf{X}|}\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} \\ &= \mathbf{X}^{-1'}, \text{ from (9.35), for } \mathbf{X} \text{ with functionally} \\ &\qquad \text{independent elements} \end{aligned} \tag{9.37}$$

$$= 2\mathbf{X}^{-1} - \text{diag}(\mathbf{X}^{-1}), \ \text{ from (9.36), for symmetric } \mathbf{X}. \tag{9.38}$$

Results (9.28) and (9.37) are well known, but (9.29) and (9.38) are not. [see Henderson and Searle (1981a)].

Another useful derivative is $\partial \log |X|/\partial y$. For X having functionally independent elements, it is

$$\frac{\partial}{\partial y}\log |X| = \frac{1}{|X|}\frac{\partial |X|}{\partial y} = \frac{1}{|X|}\sum_i\sum_j \frac{\partial |X|}{\partial x_{ij}}\frac{\partial x_{ij}}{\partial y}. \tag{9.39}$$

Observing that, in general, $tr(K'M) = \sum_\imath \sum_j k_{ij}m_{ij}$, gives

$$\begin{aligned}\frac{\partial}{\partial y}\log |X| &= \frac{1}{|X|}tr\left[\left(\frac{\partial |X|}{\partial X}\right)'\frac{\partial X}{\partial y}\right] = tr\left[\left(\frac{1}{|X|}\frac{\partial |X|}{\partial X}\right)'\frac{\partial X}{\partial y}\right]\\ &= tr\left(X^{-1}\frac{\partial X}{\partial y}\right), \quad \text{from (9.37).}\end{aligned} \tag{9.40}$$

When X is symmetric, differentiation is needed only with respect to the distinct elements of $\mathbf{X}$ so that, in contrast to (9.39),

$$\frac{\partial}{\partial y}\log |X| = \frac{1}{|X|}\sum\sum_{\imath\le j}\frac{\partial |X|}{\partial x_{\imath J}}\frac{\partial x_{\imath J}}{\partial y} \quad \text{for symmetric } X.$$

But this, in combination with (9.34), gives

$$\begin{aligned}\frac{\partial}{\partial y}\log |X| &= \frac{1}{|X|}\sum\sum_{\imath\le j}(2-\delta_{\imath J})|X_{ij}|\frac{\partial x_{\imath J}}{\partial y} \quad \text{for} \quad X = X'\\ &= \frac{1}{|X|}\sum_i\sum_J |X_{ij}|\frac{\partial x_{\imath J}}{\partial y} = \mathrm{tr}\left(X^{-1}\frac{\partial X}{\partial y}\right),\end{aligned}$$

which is (9.40); that is, (9.40) applies for symmetric and nonsymmetric X.

9.5.5 Jacobians

When y is a vector of n differentiable functions of the n elements of x, of such a nature that the transformation $x \to y$ is 1-to-1, then the matrix

$$J_{x\to y} = \left(\frac{\partial x}{\partial y}\right)' = \left\{\frac{\partial x_i}{\partial y_J}\right\} \quad \text{for} \quad i, j = 1, \ldots, n \tag{9.41}$$

is the *Jacobian matrix* of the transformation from x to y. For example, for the linear transformation, $y = Ax$, the 1-to-1 property ensures existence of A^{-1} and so $x = A^{-1}y$. Hence, using (9.15), the Jacobian matrix (9.41) for

$$y = Ax$$

is

$$\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}} = \left[\frac{\partial(\boldsymbol{A}^{-1}\boldsymbol{y})}{\partial\boldsymbol{y}}\right]' = (\boldsymbol{A}^{-1\prime})' = \boldsymbol{A}^{-1}. \tag{9.42}$$

Determinants of Jacobian matrices are required when making a change of variables in integral calculus. Suppose in an integral involving x's, a 1-to-1 transformation to y's is to be made. The integral is transformed by substituting for the x's in terms of the y's and by replacing the differentials $dx_1\, dx_2\, dx_3 \cdots dx_n$ (denoted by $d\boldsymbol{x}$) by $\|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\| dy_1\, dy_2\, dy_3 \cdots dy_n$, where $\|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\|$ is the absolute value of the determinant of $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}$. In brief, for the transformation

$$\boldsymbol{x} \to \boldsymbol{y} \text{ make the replacement } d\boldsymbol{x} \to \|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\| d\boldsymbol{y}. \tag{9.43}$$

Although $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}$ is the Jacobian matrix, $\|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\|$ is known as the *Jacobian* of the transformation.

Because the transformation in (9.43) is from $\boldsymbol{x}$ to $\boldsymbol{y}$, elements of $\boldsymbol{y}$ will usually be expressed as functions of those of $\boldsymbol{x}$ (rather than vice versa). Hence derivation of $\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}$ with elements $\partial y_i/\partial x_j$ will be easier to obtain than $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}$ with elements $\partial x_i/\partial y_j$. This presents no problem, though, because it can be shown that $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}} = \boldsymbol{I}$ and so we always have

$$\|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\| = 1/\|\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}\|. \tag{9.44}$$

Hence, whichever of $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}$ and $\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}$ is most easily obtainable from the transformation can be used in (9.43), either $\|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\|$ or $1/\|\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}\|$ of (9.44). This being so, relying upon memory for using (9.43) can raise a doubt: "Is it $\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}$ or $\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}$?" This is especially important when symbols other than $\boldsymbol{x}$ and $\boldsymbol{y}$ are being used. Fortunately, a useful mnemonic[2] is available for those who like such an aid. Denote the variables by "o" for "old" and "n" for "new," where the transformation is thought of as being from "old" variables to "new" variables. Then in (9.43) the Jacobian matrix is $\boldsymbol{J}_{o\to n}$ where the subscripts are "*on*": in contrast, the wrong matrix $\boldsymbol{J}_{n\to o}$ has subscripts "*no*"—and it is *not* "no"; that is, it is not $\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}$.

Example 9.5 *For the transformation from x_1, x_2 to y_1, y_2 represented by*

$$\left.\begin{aligned} y_1 &= e^{x_1} - x_2 \\ y_2 &= x_2 \end{aligned}\right\} \quad \textit{we have} \quad \|\boldsymbol{J}_{\boldsymbol{y}\to\boldsymbol{x}}\| = \left\|\begin{matrix} e^{x_1} & -1 \\ 0 & 1 \end{matrix}\right\| = e^{x_1} = y_1 + y_2.$$

The inverse relationship is, for $r = 1/(y_1 + y_2)$,

$$\left.\begin{aligned} x_1 &= \log(y_1 + y_2) \\ x_2 &= y_2 \end{aligned}\right\} \quad \textit{with} \quad \|\boldsymbol{J}_{\boldsymbol{x}\to\boldsymbol{y}}\| = \left\|\begin{matrix} r & r \\ 0 & 1 \end{matrix}\right\| = 1/(y_1 + y_2).$$

[2] We are indebted to Daniel L. Solomon for this.

TABLE 9.1 Jacobians for Certain Matrix Transformations

Transformation: $\mathbf{X} \to \mathbf{Y}$	Jacobian: $\|\mathbf{J}_{\mathbf{X}\to\mathbf{Y}}\|$ — Matrices Having Functionally Independent Elements	Symmetric Matrices $\mathbf{X} = \mathbf{X}'$ and $\mathbf{Y} = \mathbf{Y}'$
$\mathbf{Y}_{p\times q} = \mathbf{A}\mathbf{X}_{p\times q}\mathbf{B}$	$\lvert\mathbf{A}\rvert^{-q}\lvert\mathbf{B}\rvert^{-p}$	Not applicable
$\mathbf{Y}_{p\times p} = \mathbf{A}\mathbf{X}_{p\times p}\mathbf{A}'$	$\lvert\mathbf{A}\rvert^{-2p}$	$\lvert\mathbf{A}\rvert^{-(p+1)}$
$\mathbf{Y}_{n\times n} = \mathbf{X}^{-1}$	$\lvert\mathbf{X}\rvert^{2n}$	$\lvert\mathbf{X}\rvert^{n+1}$
$\mathbf{Y}_{n\times n} = \mathbf{X}^{k}$	$\left(\prod_{i=1}^{n}\prod_{j=1}^{n}\sum_{r=1}^{k}\lambda_i^{k-r}\lambda_j^{r-1}\right)^{-1}$	$\left(\prod_{i=1}^{n}\prod_{j\geq i}^{n}\sum_{r=1}^{k}\lambda_i^{k-r}\lambda_j^{r-1}\right)^{-1}$

Note: All determinants are taken positively, and in the last line $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalue of **X**; symmetric **X** differs from non-symmetric only in having $j \geq i$.

Sometimes variables are arrayed not as vectors but as matrices, in which case, putting the matrices in vector form (by the operations of Section 9.4) enables (9.41) and (9.43) to be used for deriving Jacobians. In general

$$J_{X\to Y} = \frac{\partial\,\mathrm{vec}\,\boldsymbol{X}}{\partial\,\mathrm{vec}\,\boldsymbol{Y}}$$

and

$$J_{X\to Y} = \frac{\partial\,\mathrm{vech}\,\boldsymbol{X}}{\partial\,\mathrm{vech}\,\boldsymbol{Y}} \text{ for } \boldsymbol{X} \text{ and } \boldsymbol{Y} \text{ symmetric.}$$

These techniques provide straightforward derivation of results such as those shown in Table 9.1. Henderson and Searle (1981a) give details and discuss earlier derivations.

9.5.6 Aitken's Integral

For positive definite $\boldsymbol{A}$, the result

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}\,dx_1\cdots dx_n = (2\pi)^{\frac{n}{2}}|\boldsymbol{A}|^{-\frac{1}{2}} \tag{9.45}$$

is known as Aitken's integral. It is established by noting that because $\boldsymbol{A}$ is positive definite, there exists a nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{I}$. Letting $\boldsymbol{y} = \boldsymbol{P}^{-1}\boldsymbol{x}$, the Jacobian for which is $\|\boldsymbol{P}\| = |\boldsymbol{A}|^{-\frac{1}{2}}$, gives

$$\begin{aligned}
&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}\,dx_1\cdots dx_n \\
&\quad = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}\boldsymbol{y}}\,dy_1\cdots dy_n\,|\boldsymbol{A}|^{-\frac{1}{2}} \\
&\quad = \prod_{i=1}^{n}\left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2}\,dy_i\right]|\boldsymbol{A}|^{-\frac{1}{2}} \\
&\quad = (2\pi)^{\frac{n}{2}}|\boldsymbol{A}|^{-\frac{1}{2}}.
\end{aligned}$$

9.5.7 Hessians

Another matrix of partial differentials is the Hessian. If θ is a function of n variables $x_1, x_2, \ldots, x_n$, the matrix of second-order partial derivatives of θ with respect to the x's is called a Hessian: for $i, j = 1, 2, \ldots, n$,

$$\mathbf{H} = \left\{ \frac{\partial^2 \theta}{\partial x_i \partial x_j} \right\} = \frac{\partial^2 \theta}{\partial \mathbf{x} \partial \mathbf{x}'}.$$

Example 9.6 *One of the properties of maximum likelihood in statistics is that, under certain conditions [see, for example, Mood et al. (1974)], the variance-covariance matrix of a set of simultaneously estimated parameters is the inverse of minus the expected value of the Hessian of the likelihood with respect to those parameters.*

9.6 MATRICES WITH ELEMENTS THAT ARE COMPLEX NUMBERS

Knowledge of complex numbers is not a prerequisite for reading this book, which deals primarily with real matrices, that is, having elements that are real numbers. However, as with other topics that have been touched on lightly in order to introduce the reader to concepts and vocabulary that may be encountered elsewhere, so here–we briefly give some definitions of matrices having elements that are complex numbers.

The starting point is the square root of a negative number. For $\sqrt{-1}$ the symbol i is used so that, for example, $\sqrt{-6} = i\sqrt{6}$ and $\sqrt{-81} = 9i$. Numbers that are multiples of i are called imaginary numbers; those not involving imaginary numbers are real. The sum of a real and an imaginary number is a complex number; that is, $3 + 7i$ and $-2 + i\sqrt{11}$ are complex numbers. In general, for a and b being real, $a + ib$ is complex. The number $a - ib$ is called the complex conjugate of $a + ib$ and the pair of numbers $a + ib$ and $a - ib$ are called a conjugate pair.

An example of a matrix having complex numbers for elements is

$$\mathbf{M} = \begin{bmatrix} 3+2i & 4-5i \\ -6i & 9+i\sqrt{2} \\ 7-i & 6+7i \end{bmatrix}; \quad \text{and} \quad \overline{\mathbf{M}} = \begin{bmatrix} 3-2i & 4+5i \\ 6i & 9-i\sqrt{2} \\ 7+i & 6-7i \end{bmatrix}$$

is the *complex conjugate of* $\mathbf{M}$. It is the matrix whose elements are the complex conjugates of the elements of $\mathbf{M}$. If, for $\mathbf{A}$ and $\mathbf{B}$ being real, $\mathbf{M} = \mathbf{A} + i\mathbf{B}$, then $\overline{\mathbf{M}} = \mathbf{A} - i\mathbf{B}$, analogous to scalar complex numbers. *A Hermitian matrix* is one in which the transpose of the complex conjugate equals the matrix itself: $\overline{\mathbf{M}}' = \mathbf{M}$, in which case the symbol $\mathbf{M}^*$ is often used. For example, if

$$\mathbf{M}^* = \begin{bmatrix} 3 & 7+8i \\ 7-8i & 4 \end{bmatrix},$$

$$\overline{\mathbf{M}}^{*\prime} = \begin{bmatrix} 3 & 7-8i \\ 7+8i & 4 \end{bmatrix}' = \begin{bmatrix} 3 & 7+8i \\ 7-8i & 4 \end{bmatrix} = \mathbf{M}^*.$$

The diagonal elements of a Hermitian matrix are real. When $\overline{\boldsymbol{M}}'\boldsymbol{M} = \boldsymbol{I}$, the matrix $\boldsymbol{M}$ is said to be a *unitary* matrix.

Clearly, when $\boldsymbol{B}$ is null, matrices of the form $\boldsymbol{A} + i\boldsymbol{B}$ become real, if Hermitian they become symmetric, and if unitary they become orthogonal; that is, being Hermitian is the counterpart of being symmetric, as is unitary of orthogonal. Indeed, symmetric matrices are a subset of Hermitian matrices, as are orthogonal matrices of unitary ones.

9.7 MATRIX INEQUALITIES

This section presents several inequalities that pertain to matrices, some of which represent matrix analogues of well-known inequalities such as the Cauchy–Schwarz inequality. These inequalities are useful in many aspects of linear models and multivariate statistics. The objective here is to provide an exposition of such inequalities without delving into the process of proving most of them. The reader is referred to other sources where proofs can be found.

Lemma 9.1 *Let $\boldsymbol{A}$ be a symmetric matrix or order $n \times n$. Then,*

(a) *The matrix $[e_{max}(\boldsymbol{A})]\boldsymbol{I}_n - \boldsymbol{A}$ is positive semidefinite,*
(b) *The matrix $\boldsymbol{A} - [e_{min}(\boldsymbol{A})]\boldsymbol{I}_n$ is positive semidefinite,*

where $e_{min}(\boldsymbol{A})$ and $e_{max}(\boldsymbol{A})$ are, respectively, the smallest and largest eigenvalues of $\boldsymbol{A}$.

Proof. (a) The proof follows by applying the spectral decomposition theorem to $\boldsymbol{A}$: $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix whose columns are corresponding eigenvectors (see Theorem 6.2). Thus, for all $\boldsymbol{x}$, $\boldsymbol{x}'\{[e_{max}(\boldsymbol{A})]\boldsymbol{I}_n - \boldsymbol{A}\}\boldsymbol{x} = \boldsymbol{y}'\{[e_{max}(\boldsymbol{A})]\boldsymbol{I}_n - \boldsymbol{\Lambda}\}\boldsymbol{y} \geq 0$, where $\boldsymbol{y} = \boldsymbol{P}'\boldsymbol{x}$. The proof of (b) is similar. ■

Corollary 9.2 *If $\boldsymbol{A}$ is a symmetric matrix, then*

$$e_{min}(\boldsymbol{A}) \leq \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \leq e_{max}(\boldsymbol{A}).$$

Equalities hold for both the lower and upper bounds if $\boldsymbol{x}$ is chosen an eigenvector of $\boldsymbol{A}$ corresponding to $e_{min}(\boldsymbol{A})$ and $e_{max}(\boldsymbol{A})$, respectively. It follows that

$$\inf_{\boldsymbol{x}\neq\boldsymbol{0}} \left[\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\right] = e_{min}(\boldsymbol{A})$$
$$\sup_{\boldsymbol{x}\neq\boldsymbol{0}} \left[\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\right] = e_{max}(\boldsymbol{A}).$$

The ratio $\dfrac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}$ is known as *Rayleigh's quotient*.

Corollary 9.3 *If $\boldsymbol{A}$ is a symmetric matrix and $\boldsymbol{B}$ is a positive definite matrix, both of order $n \times n$, then*

$$e_{min}(\boldsymbol{B}^{-1}\boldsymbol{A}) \leq \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}} \leq e_{max}(\boldsymbol{B}^{-1}\boldsymbol{A}).$$

Proof. This follows from Corollary 9.2 by applying the transformation $\boldsymbol{x} = \boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{y}$ to the ratio $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ which takes the form

$$\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}} = \frac{\boldsymbol{y}'\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}.$$

Hence, by Corollary 9.2,

$$e_{min}(\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}) \leq \frac{\boldsymbol{y}'\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \leq e_{max}(\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}).$$

The proof now follows from noting that the eigenvalues of $\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}$ are the same as the eigenvalues of $\boldsymbol{B}^{-1}\boldsymbol{A}$ (see Theorem 6.12, part c). Note also that the eigenvalues of $\boldsymbol{B}^{-1}\boldsymbol{A}$ are real since they are the same as those of the matrix $\boldsymbol{C} = \boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{-\frac{1}{2}}$, which is symmetric (see Lemma 6.1). Here, $\boldsymbol{B}^{-\frac{1}{2}}$ is a matrix defined as follows: by the spectral decomposition theorem, $\boldsymbol{B}$ can be written as $\boldsymbol{B} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{B}$ and $\boldsymbol{P}$ is an orthogonal matrix whose columns are the corresponding eigenvectors of $\boldsymbol{B}$. Then, $\boldsymbol{B}^{-\frac{1}{2}}$ is defined as

$$\boldsymbol{B}^{-\frac{1}{2}} = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}',$$

where

$$\boldsymbol{\Lambda}^{-\frac{1}{2}} = \text{diag}(\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \ldots, \lambda_n^{-\frac{1}{2}}).$$

■

Theorem 9.6 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of the same order. Then,*

$$[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \leq [tr(\boldsymbol{A}'\boldsymbol{A})][tr(\boldsymbol{B}'\boldsymbol{B})].$$

Equality holds if and only if one of these two matrices is a scalar multiple of the other.

[Magnus and Neudecker (1988), Theorem 2, p. 201]. This is a matrix analogue of the *Cauchy–Schwarz* inequality: If $\boldsymbol{a}$ and $\boldsymbol{b}$ are two vectors having the same number of elements, then $(\boldsymbol{a}'\boldsymbol{b})^2 \leq (\boldsymbol{a}'\boldsymbol{a})(\boldsymbol{b}'\boldsymbol{b})$; equality holds if and only if one vector is a scalar product of the other.

Another version of the Cauchy–Schwartz inequality for matrices is given by the next theorem.

Theorem 9.7 *Let $\boldsymbol{A}$ be a positive definite matrix. Then,*

$$(\boldsymbol{u}'\boldsymbol{v})^2 \leq (\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u})(\boldsymbol{v}'\boldsymbol{A}^{-1}\boldsymbol{v}).$$

Equality holds if and only if $\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u}$ for some real number λ.
In particular,

$$(\boldsymbol{u}'\boldsymbol{u})^2 \leq (\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u})(\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}). \tag{9.46}$$

Equality is attained if and only if $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$.

[Puntanen et al. (2011), Section 20.1, p. 417].

Theorem 9.8 *Let $\boldsymbol{A} = (a_{ij})$ be a positive definite matrix of order $n \times n$. Then,*

$$|\boldsymbol{A}| \leq \prod_{i=1}^{n} a_{ii}.$$

Equality holds if and only if $\boldsymbol{A}$ is diagonal.

[Magnus and Neudecker (1988), Theorem 28, p. 23].

Corollary 9.4 *For any real matrix $\boldsymbol{A} = (a_{ij})$ of order $n \times n$,*

$$|\boldsymbol{A}|^2 \leq \prod_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}^2\right).$$

Equality holds if and only if $\boldsymbol{A}'\boldsymbol{A}$ is a diagonal matrix or $\boldsymbol{A}$ has a zero row.

Proof. If $\boldsymbol{A}$ is singular, the result is trivially true; assume therefore that $\boldsymbol{A}$ is nonsingular. Then, $\boldsymbol{A}'\boldsymbol{A}$ is positive definite. Applying Theorem 9.8 to $\boldsymbol{A}'\boldsymbol{A}$, we get

$$|\boldsymbol{A}|^2 = |\boldsymbol{A}'\boldsymbol{A}| \leq \prod_{j=1}^{n} \boldsymbol{c}_j'\boldsymbol{c}_j,$$

since $\boldsymbol{c}_j'\boldsymbol{c}_j$ is the jth diagonal element of $\boldsymbol{A}'\boldsymbol{A}$, where $\boldsymbol{c}_j$ is the jth column of $\boldsymbol{A}, j = 1, 2, \ldots, n$. Thus, $\boldsymbol{c}_j'\boldsymbol{c}_j = \sum_{i=1}^{n} a_{ij}^2$. Hence,

$$|\boldsymbol{A}|^2 \leq \prod_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}^2\right).$$

■

This inequality is called *Hadamard's inequality*.

It can also be shown that

$$|A|^2 \leq \prod_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right). \tag{9.47}$$

Equality holds if and only if $\boldsymbol{AA'}$ is a diagonal matrix or at least one of the rows of $\boldsymbol{A}$ is $\boldsymbol{0}'$.

Theorem 9.9 *For any two positive semidefinite matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, of order $n \times n$,*

$$|\boldsymbol{A}+\boldsymbol{B}|^{\frac{1}{n}} \geq |\boldsymbol{A}|^{\frac{1}{n}} + |\boldsymbol{B}|^{\frac{1}{n}}.$$

Equality holds if and only if $|\boldsymbol{A}+\boldsymbol{B}| = 0$, or $\boldsymbol{A} = \alpha\boldsymbol{B}$ for some $\alpha > 0$.

This is called *Minkowski's determinant inequality* [see Magnus and Neudecker (1988, Theorem 28, p. 227)].

Theorem 9.10 *If $\boldsymbol{A}$ is a positive semidefinite matrix and $\boldsymbol{B}$ is a positive definite matrix, both of order $n \times n$, then for any $i (i = 1, 2, \ldots, n)$,*

$$e_i(\boldsymbol{A})e_{min}(\boldsymbol{B}) \leq e_i(\boldsymbol{AB}) \leq e_i(\boldsymbol{A})e_{max}(\boldsymbol{B}).$$

In particular,

$$e_{min}(\boldsymbol{A})e_{min}(\boldsymbol{B}) \leq e_i(\boldsymbol{AB}) \leq e_{max}(\boldsymbol{A})e_{max}(\boldsymbol{B}).$$

Furthermore, if $\boldsymbol{A}$ is positive definite, then for any i $(i = 1, 2, \ldots, n)$,

$$\frac{e_i^2(\boldsymbol{AB})}{e_{max}(\boldsymbol{A})e_{max}(\boldsymbol{B})} \leq e_i(\boldsymbol{A})e_i(\boldsymbol{B}) \leq \frac{e_i^2(\boldsymbol{AB})}{e_{min}(\boldsymbol{A})e_{min}(\boldsymbol{B})}.$$

[Anderson and Gupta (1963), Corrolary 2.2.1].

Theorem 9.11 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices of order $n \times n$. Then,*

(a) *$e_i(\boldsymbol{A}) \leq e_i(\boldsymbol{A}+\boldsymbol{B})$, $i = 1, 2, \ldots, n$, if $\boldsymbol{B}$ is non-negative definite.*
(b) *$e_i(\boldsymbol{A}) < e_i(\boldsymbol{A}+\boldsymbol{B})$, $i = 1, 2, \ldots, n$, if $\boldsymbol{B}$ is positive definite.*

[Bellman (1997), Theorem 3, p. 117].

Theorem 9.12 *Let $\boldsymbol{A}$ be a positive definite matrix and $\boldsymbol{B}$ be a positive semidefinite matrix, both of order $n \times n$. Then,*

$$|\boldsymbol{A}| \leq |\boldsymbol{A}+\boldsymbol{B}|.$$

Equality holds if and only if $\boldsymbol{B} = \boldsymbol{0}$.

[Magnus and Neudecker (1988), Theorem 22, p. 21].

Theorem 9.13 *Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$, and let $\| \mathbf{A} \|_2$ denote its Euclidean norm, that is,*

$$\| \mathbf{A} \|_2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}}.$$

Then,

$$\sum_{i=1}^{n} e_i^2(\mathbf{A}) = \| \mathbf{A} \|_2^2 .$$

[This is known as *Schur's Theorem*. See Lancaster (1969), Theorem 7.3.1].

Corollary 9.5

$$| e_{max}(\mathbf{A}) | \leq n [\max_{i,j} | a_{ij} |].$$

Proof. We have that $\| \mathbf{A} \|_2 \leq n [\max_{i,j} | a_{ij} |]$. Apply now Theorem 9.13 to get the desired result. ■

Theorem 9.14 *Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$, and let m and s be scalars defined as*

$$m = \frac{tr(\mathbf{A})}{n}, \qquad s = \left[\frac{tr(\mathbf{A}^2)}{n} - m^2 \right]^{\frac{1}{2}}.$$

Then,

$$m - s(n-1)^{\frac{1}{2}} \leq e_{min}(\mathbf{A}) \leq m - \frac{s}{(n-1)^{\frac{1}{2}}}$$

$$m + \frac{s}{(n-1)^{\frac{1}{2}}} \leq e_{max}(\mathbf{A}) \leq m + s(n-1)^{\frac{1}{2}}$$

$$e_{max}(\mathbf{A}) - e_{min}(\mathbf{A}) \leq s(2n)^{\frac{1}{2}}.$$

[Wolkowicz and Styan (1980), Theorems 2.1 and 2.5].

REFERENCES

Amghibech, S., Chachulska, J., and Matysiak, W. (2008). A matrix inequality: 11296. *Am. Math. Month.*, 115, p. 671.

Anderson, T. W. and Gupta, S. D. (1963). Some inequalities on characteristic roots of matrices. *Biometrika*, 50, 522–524.

Bellman, R. (1997). *Introduction to Matrix Analysis*, 2nd ed. SIAM, Philadelphia.

Deemer, W. L. and Olkin, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. Based on lectures of P. l. Hsu at the University of North Carolina, 1947. *Biometrika*, 38, 345–367.

Dwyer, P. S. (1967). Some applications of matrix derivations in multivariate analysis. *J. Am. Stat. Assoc.*, 62, 607–625.

Golub, G. H. and Van Loan, C. F. (1983). *Matrix Computations*. John Hopkins University Press, Baltimore, MD.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, CA.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*, 2nd ed. University Press, Cambridge, UK.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28, 476–490.

Henderson, H. V. and Searle, S. R. (1979). Vec and Vech operators for matrices with some uses in Jacobians and multivariate statistics. *Can. J. Stat.*, 7, 65–81.

Henderson, H. V. and Searle S. R. (1981a). On deriving the inverse of a sum of matrices. *SIAM Rev.*, 23, 53–60.

Henderson, H. V. and Searle, S. R. (1981b). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra*, 9, 271–288.

Lancaster, P. (1969). *Theory of Matrices*. Academic Press, New York.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*, 3rd ed. McGraw-Hill, New York.

Nel, D. G. (1980). On matrix differentiation in statistics. *South African J. Stat.*, 14, 137–193.

Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Am. Stat. Assoc.*, 64, 953–963.

Puntanen, S., Styan, G. P. H., and Isotalo, J. (2011). *Matrix Tricks for Linear Satistical Models*. Springer, New York.

Rogers, G. S. (1980). *Matrix Derivatives*. Dekker, New York.

Wolkowicz, H. and Styan, G. P. H. (1979). Extensions of Samuelson's inequality. *Am. Stat.*, 33, 143–144.

Wolkowicz, H. and Styan, G. P. H. (1980). Bounds for eigenvalues using traces. *Linear Algebra Appl.*, 29, 471–506.

EXERCISES

9.1 Show that with respect to x_1 and x_2, the Jacobian of

$$y_1 = 6x_1^2 x_2 + 2x_1 x_2 + x_2^2 \quad \text{and} \quad y_2 = 2x_1^3 + x_1^2 + 2x_1 x_2$$

is the same as the Hessian of

$$y = 2x_1^3 x_2 + x_1^2 x_2 + x_1 x_2^2.$$

9.2 Express the quadratic form

$$7x_1^2 + 4x_1x_2 - 5x_2^2 - 6x_2x_3 + 3x_3^2 + 6x_1x_3$$

in the form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is symmetric and verify (9.17).

9.3 Find the value of $\boldsymbol{\beta}$ that maximizes $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{t}\boldsymbol{t}'\boldsymbol{X}\boldsymbol{\beta}/\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$; and find that maximum value. The matrix $\boldsymbol{X}$ is of order $n \times p$ and rank p $(n \geq p)$.

9.4 Prove inequality (9.47).

9.5 For $\boldsymbol{A}$ being non-negative definite, prove that $(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y})$. Applications of this to statistics are given by Wolkowicz and Styan (1979). (Hint: Use the Cauchy inequality.)

9.6 Verify: $tr(\boldsymbol{A}\boldsymbol{X}) = (\text{vec}\boldsymbol{A}')'\text{vec}\boldsymbol{X} = (\text{vec}\boldsymbol{X}')'\text{vec}\boldsymbol{A}$.

9.7 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices. Show that

$$tr[(\boldsymbol{A}\boldsymbol{B})^2] \leq tr(\boldsymbol{A}^2\boldsymbol{B}^2)$$

Under what condition is equality valid? (Hint: Find the trace of $\boldsymbol{C}\boldsymbol{C}'$, where $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}$.)

9.8 Find the Hessian matrix of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a symmetric matrix.

9.9 Find the derivative of $\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}$ with respect to the matrix $\boldsymbol{W}$.

9.10 Suppose that $\boldsymbol{z} = (z_1, z_2, \ldots, z_p)'$ is a function of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, which is a function of $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)'$. Show that the $m \times p$ matrix $\dfrac{\partial \boldsymbol{z}'(\boldsymbol{x})}{\partial \boldsymbol{x}}$ can be expressed as:

$$\frac{\partial \boldsymbol{z}'(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{y}'}{\partial \boldsymbol{x}}\right)\left(\frac{\partial \boldsymbol{z}'}{\partial \boldsymbol{y}}\right).$$

This is called the *chain rule* for vector functions. Note that this rule proceeds from right to left. Thus, if $\boldsymbol{u}$ is a function of $\boldsymbol{z}$, which is a function of $\boldsymbol{y}$, which is a function of $\boldsymbol{x}$, then

$$\frac{\partial \boldsymbol{u}'}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{y}'}{\partial \boldsymbol{x}}\right)\left(\frac{\partial \boldsymbol{z}'}{\partial \boldsymbol{y}}\right)\left(\frac{\partial \boldsymbol{u}'}{\partial \boldsymbol{z}}\right).$$

9.11 Let $\boldsymbol{A}$ be a positive semidefinite matrix of order $n \times n$. Show that

$$tr(\boldsymbol{A}^2) \leq tr(\boldsymbol{A})e_{max}(\boldsymbol{A}),$$

where $e_{max}(\boldsymbol{A})$ is the largest eigenvalue of $\boldsymbol{A}$. Under what condition does equality hold?

9.12 Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be matrices of orders $p \times q$ and $m \times n$, respectively. Show that

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} = \sum_{i=1}^{p} \sum_{j=1}^{q} \boldsymbol{E}_{ij} \otimes \frac{\partial \boldsymbol{Y}}{\partial x_{ij}},$$

where E_{ij} is a matrix of order $p \times q$ whose elements are equal to 0, except for the (i, j)th element, which is equal to 1.

9.13 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices of order $m \times m$. Show that

$$tr(\boldsymbol{AB}) \leq \frac{1}{2} tr(\boldsymbol{A}^2 + \boldsymbol{B}^2).$$

9.14 Let $\boldsymbol{A}$ and $\boldsymbol{A}$ be square matrices of the same order. Show that

$$tr(\boldsymbol{A} \otimes \boldsymbol{B}) \leq \frac{1}{2} tr(\boldsymbol{A} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{B}).$$

9.15 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices of order $n \times n$. The matrix $\boldsymbol{A}$ is assumed to be non-negative definite. Show that

$$e_{min}(\boldsymbol{B}) tr(\boldsymbol{A}) \leq tr(\boldsymbol{AB}) \leq e_{max}(\boldsymbol{B}) \mathrm{tr}(\boldsymbol{A}).$$

9.16 Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote its eigenvalues. Suppose that the real-valued function $f(x)$ can be represented as a power series of the form

$$f(x) = \sum_{i=0}^{\infty} a_i x^i,$$

which is defined on open interval that contains the eigenvalues of $\boldsymbol{A}$. By the spectral decomposition theorem (Theorem 6.2), the matrix $\boldsymbol{A}$ can be represented as

$$\boldsymbol{A} = \boldsymbol{P\Lambda P}',$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\boldsymbol{P}$ is an orthogonal matrix whose columns are the corresponding eigenvectors of $\boldsymbol{A}$. The matrix function, $f(\boldsymbol{A})$ can be defined as [see Golub and Van Loan (1983), Corollary 11.1–2, p. 382]

$$f(\boldsymbol{A}) = \boldsymbol{P} \mathrm{diag}[f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)] \boldsymbol{P}'.$$

Show that $f(\boldsymbol{A})$ can be represented as the sum of a convergent power series of the form

$$f(\boldsymbol{A}) = \sum_{i=0}^{\infty} a_i \boldsymbol{A}^i,$$

if $\sum_{i=0}^{\infty} a_i\lambda_1^i$, $\sum_{i=0}^{\infty} a_i\lambda_2^i$, ..., $\sum_{i=0}^{\infty} a_i\lambda_n^i$ are all convergent power series and represent $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$, respectively. (Hint: See Golub and Van Loan, 1983, Theorem 11.2-3, p. 390.)

9.17 Apply the result in Problem 16 to find a power series representation of $sin(\boldsymbol{A})$, where $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$.

9.18 Let $\boldsymbol{X}$ be a symmetric matrix of order $n \times n$ of independent elements, except that $x_{ij} = x_{ji}$, $i \neq j$, $i, j = 1, 2, \ldots, n$. Let $\boldsymbol{a}$ be a constant vector of n elements. Show that

$$\frac{\partial(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a})}{\partial \boldsymbol{X}} = 2\,\boldsymbol{a}\boldsymbol{a}' - \boldsymbol{D}_{\boldsymbol{a}\boldsymbol{a}'},$$

where $\boldsymbol{D}_{\boldsymbol{a}\boldsymbol{a}'}$ is a diagonal matrix of order $n \times n$ whose diagonal elements are the same as those of $\boldsymbol{a}\boldsymbol{a}'$. (Hint: See Graybill, 1983, Theorem 10.8.4, p. 354.)

9.19 Consider equality (9.31). Show that if $\dfrac{\partial(\boldsymbol{Q}\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{0}$, then $\boldsymbol{Q}$ must be equal to a zero matrix.

9.20 Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote its eigenvalues such that $|\lambda_i| < 1$ for $i = 1, 2, \ldots, n$. Show that the power series in $\boldsymbol{A}$, $\sum_{j=0}^{\infty} \boldsymbol{A}^j$, converges to $(\boldsymbol{I}_n - \boldsymbol{A})^{-1}$.

9.21 Let $\boldsymbol{A} = (a_{ij})$ be a symmetric positive definite matrix of order $n \times n$. Let $\boldsymbol{A}_j$ be the matrix of order $j \times j$ in the upper left corner of $\boldsymbol{A}$ (that is, the leading principal submatrix of $\boldsymbol{A}$ of order $j \times j$, $j = 1, 2, \ldots, n$). Show that

$$\sum_{j=1}^{n} |\boldsymbol{A}_j|^{1/j} \leq \left(1 + \frac{1}{n}\right)^n tr(\boldsymbol{A}).$$

(Hint: Use Theorem 9.8 and *Carleman's inequality* that states

$$\sum_{j=1}^{n} (x_1 x_2 \ldots x_j)^{1/j} \leq \left(1 + \frac{1}{n}\right)^n \sum_{j=1}^{n} x_j,$$

for $x_j \geq 0$, $j = 1, 2, \ldots, n$ [see Hardy et al., 1952, p. 249, and Amghibech et al. 2008, p. 671].)

Part II

Applications of Matrices in Statistics

In this part we begin an exhibition of certain areas of statistics that utilize matrix techniques. It is demonstrated that the development and understanding of several statistical methods are greatly enhanced by the effective use of matrix algebra.

Part II consists of the following chapters:

Chapter 10: Multivariate Statistics.
This chapter includes a study of the multivariate normal distribution and related distributions, its moment generating function, and associated quadratic forms. Results pertaining to quadratic forms, such as their distribution and independence under normal theory, are discussed.

Chapter 11: Matrix Algebra of Full-Rank Linear Models.
This chapter covers estimation by the method of least squares and its properties, analysis of variance, the Gauss-Markov theorem, generalized least squares estimation, testing of linear hypotheses, and confidence regions and intervals.

Chapter 12: Less-Than-Full-Rank Linear Models.
Emphasis in this chapter is placed on the concept of estimability of linear functions and testable hypotheses, in addition to a study of some particular models such as the one-way and two-way classification models.

Chapter 13: Analysis of Balanced Linear Models Using Direct Product of Matrices.
Properties associated with such models are discussed which result in setting up a general treatment of their analyses.

Chapter 14: Multiresponse Models.
This chapter includes a study of linear multiresponse models and their analyses, including the testing of lack of fit.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

10

Multivariate Distributions and Quadratic Forms

Matrix algebra plays a central role in facilitating the development and presentation of many statistical methods. A few examples from a multitude of possibilities are presented in this chapter and the remaining chapters in Part II.

Notation Thus far we have maintained the conventional mathematics notation of using capital letters for matrices and lowercase for vectors. This will be continued, despite its clash with the statistics convention of using capital letters for random variables and lowercase for realized values thereof. For example, the use of $\boldsymbol{X}$ for a matrix and $\boldsymbol{x}$ for a vector will be maintained, whereas the notation $\boldsymbol{X} = [X_1\ X_2\ \ldots\ X_k]'$ for a vector of random variables $X_1, X_2, \ldots, X_k$ will not be used. The matrix use of $\boldsymbol{X}$ is retained by making $\boldsymbol{x} = \{x_i\}$ do double duty: on some occasions $\boldsymbol{x}$ will represent random variables and on other occasions it will represent realized values. Maintaining this distinction is easier than having $\boldsymbol{X}$ sometimes be a vector when it is usually a matrix.

The operation of taking expected values will be represented by E. Thus $E(\boldsymbol{x})$ is the vector of the expected value of each element of $\boldsymbol{x}$:

$$E(\boldsymbol{x}) = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_k) \end{bmatrix}.$$

Similarly, $E(\boldsymbol{X})$ is the matrix of the expected value of each element of $\boldsymbol{X}$.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

10.1 VARIANCE-COVARIANCE MATRICES

Suppose that $x_1, x_2, \ldots, x_k$ are k random variables with means $\mu_1, \mu_2, \ldots, \mu_k$, variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ and covariances $\sigma_{12}, \sigma_{13}, \ldots, \sigma_{k-1,k}$. On representing the random variables and their means by the vectors,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \text{we have} \quad E(\boldsymbol{x}) = \boldsymbol{\mu}; \tag{10.1}$$

and assembling the variances and covariances in a matrix gives

$$\text{var}(\boldsymbol{x}) = \boldsymbol{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & & \vdots \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_k^2 \end{bmatrix}. \tag{10.2}$$

This is the *variance–covariance matrix*, or *dispersion matrix*, of the random variables $x_1, x_2, \ldots, x_k$. It is symmetric, $\boldsymbol{V}' = \boldsymbol{V}$; its ith diagonal element is the variance of x_i and its (ij)th off-diagonal element (for $i \neq j$) is the covariance between x_i and x_j.

The familiar definitions of variance and covariance are

$$\sigma_i^2 = E(x_i - \mu_i)^2 \quad \text{and} \quad E[(x_i - \mu_i)(x_j - \mu_j)] = \sigma_{ij} \quad \text{for} \quad i \neq j,$$

respectively. They provide the means for expressing $\boldsymbol{V}$ as

$$\boldsymbol{V} = \text{var}(\boldsymbol{x}) = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})']. \tag{10.3}$$

When the means are zero, $\boldsymbol{\mu} = \boldsymbol{0}$, and (10.3) becomes

$$\boldsymbol{V} = \text{var}(\boldsymbol{x}) = E(\boldsymbol{x}\boldsymbol{x}').$$

The dispersion matrix of a linear transformation $\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x}$ of the x_i's is easily obtained: first we have the mean,

$$\text{E}(\boldsymbol{y}) = \text{E}(\boldsymbol{T}\boldsymbol{x}) = \boldsymbol{T}\,\text{E}(\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{\mu}. \tag{10.4}$$

Then the dispersion matrix of $\boldsymbol{y}$ is

$$\text{var}(\boldsymbol{y}) = \text{E}[\boldsymbol{y} - \text{E}(\boldsymbol{y})][\boldsymbol{y} - \text{E}(\boldsymbol{y})]' = \text{E}[\boldsymbol{T}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{T}'] = \boldsymbol{T}\boldsymbol{V}\boldsymbol{T}'. \tag{10.5}$$

When $\boldsymbol{T}$ is just a single row, that is, a row vector $\boldsymbol{t}'$, (10.5) is $\text{var}(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{t}'\boldsymbol{V}\boldsymbol{t}$. Because $\text{var}(\boldsymbol{t}'\boldsymbol{y})$ is a variance, it is never negative and so therefore neither is $\boldsymbol{t}'\boldsymbol{V}\boldsymbol{t}$. Thus, $\boldsymbol{t}'\boldsymbol{V}\boldsymbol{t}$ is a non-negative definite quadratic form. Hence, $\boldsymbol{V}$ is a non-negative definite matrix (see Section 5.7). This is a characteristic of all variance–covariance matrices; they are all non-negative definite. In many cases they are specifically positive definite.

10.2 CORRELATION MATRICES

Suppose ρ_{ij} is the correlation between the random variables x_i and x_j. Then, provided every σ_i^2 is nonzero,

$$\rho_{ij} = \rho_{ji} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \tag{10.6}$$

and arraying the ρ_{ij}'s in a (symmetric) matrix $\boldsymbol{R}$, with diagonal elements $\rho_{ii} = \sigma_i^2/\sigma_i^2 = 1$, gives the correlation matrix,

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{12} & 1 & \cdots & \rho_{2k} \\ \vdots & & & \vdots \\ \rho_{1k} & \rho_{2k} & \cdots & 1 \end{bmatrix}. \tag{10.7}$$

Define

$$\boldsymbol{D} = \text{diag}\{\sigma_i^2\} \quad \text{for} \quad i = 1, \ldots, k,$$

a diagonal matrix of the variances in the diagonal of $\boldsymbol{V}$ and observe that $\sqrt{\boldsymbol{D}} = \boldsymbol{D}^{\frac{1}{2}} = \text{diag}\{\sigma_i\}$. Then,

$$\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{V} \boldsymbol{D}^{-\frac{1}{2}}. \tag{10.8}$$

Example 10.1 *Suppose* $\mathbf{x}$ *represents k random variables each having variance* σ^2*, with the covariance between every pair of* $\mathbf{x}$*'s being* $\rho\sigma^2$*. Then, the variance–covariance matrix of* $\mathbf{x}$ *is, from (10.2),*

$$\boldsymbol{V} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \boldsymbol{R} \quad \textit{for} \quad \boldsymbol{R} = (1-\rho)\boldsymbol{I}_k + \rho \boldsymbol{J}_k.$$

When the $\mathbf{x}$*'s follow a normal distribution, the determinant and inverse of the correlation matrix* $\boldsymbol{R}$ *are needed for the density function. They are*

$$|\boldsymbol{R}| = (1-\rho)^{k-1}(1-\rho+k\rho)$$

and

$$\boldsymbol{R}^{-1} = \frac{1}{1-\rho}\left(\boldsymbol{I}_k - \frac{\rho}{1-\rho+k\rho}\boldsymbol{J}_k\right).$$

Since $\boldsymbol{V}$ *is positive semidefinite,* $|\boldsymbol{V}|$ *and* $|\boldsymbol{R}|$ *are non-negative, and therefore feasible values for* ρ *are restricted to* $1 \geq \rho \geq -1/(k-1)$*. If either equality holds,* $\boldsymbol{R}^{-1}$ *does not exist.*

10.3 MATRICES OF SUMS OF SQUARES AND CROSS-PRODUCTS

10.3.1 Data Matrices

We use the symbol x in this section to represent an observation. Suppose we have N samples of data on each of k variables, with x_{ij} being the ith observation on variable j. Then it is natural to array all Nk observations in what is called a *data matrix*:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} \tag{10.9}$$

$$= \{x_{ij}\} \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad j = 1, \ldots, k. \tag{10.10}$$

In this definition of $\boldsymbol{X}$ of order $N \times k$, each column represents the N observations on a particular variable and each row represents a single observation on all k variables.

10.3.2 Uncorrected Sums of Squares and Products

The uncorrected sum of squares of the observations on variable j in $\boldsymbol{X}$ is

$$\sum_{i=1}^{N} x_{ij}^2 = \boldsymbol{x}_j'\boldsymbol{x}_j,$$

where $\boldsymbol{x}_j$ is the jth column of $\boldsymbol{X}$; similarly, the corresponding sum of products of the observations on the jth and tth variables is

$$\sum_{i=1}^{N} x_{ij}x_{it} = \boldsymbol{x}_j'\boldsymbol{x}_t = \boldsymbol{x}_t'\boldsymbol{x}_j.$$

Assembling all of these sums of squares and cross-products into a matrix, which is symmetric, we have

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} \sum_{l=1}^{N} x_{l1}^2 & \sum_{i=1}^{N} x_{i1}x_{l2} & \cdots & \sum_{l=1}^{N} x_{i1}x_{ik} \\ \sum_{l=1}^{N} x_{l1}x_{l2} & \sum_{i=1}^{N} x_{l2}^2 & \cdots & \sum_{l=1}^{N} x_{l2}x_{ik} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} x_{l1}x_{ik} & \sum_{l=1}^{N} x_{l2}x_{lk} & \cdots & \sum_{i=1}^{N} x_{lk}^2 \end{bmatrix} \tag{10.11}$$

as the matrix of uncorrected sums of squares and cross-products of the x's. It is sometimes called the *uncorrected sum of squares* (and cross-products) *matrix*.

10.3.3 Means, and the Centering Matrix

The mean of the N observations on the jth variable, that is, of the elements in the jth column of $\boldsymbol{X}$, is

$$\bar{x}_{\cdot j} = \sum_{i=1}^{N} x_{ij}/N = (x_{1j} + \cdots + x_{Nj})/N = \mathbf{1}'\boldsymbol{x}_j/N = \boldsymbol{x}_j{}'\mathbf{1}/N, \tag{10.12}$$

where $\mathbf{1}$ is the summing vector (all elements are equal to one), of order N. This is the observed mean for the jth variable. And the row vector of all k of such observed means is

$$\bar{\boldsymbol{x}}' = (\bar{x}_{\cdot 1}, \bar{x}_{\cdot 2}, \ldots, \bar{x}_{\cdot k}) = \frac{1}{N}\mathbf{1}'\boldsymbol{X}. \tag{10.13}$$

Now consider each element x_{ij} of $\boldsymbol{X}$ replaced by its deviation from its column mean: for $i = 1, \ldots, N$ and $j = 1, \ldots, k$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} - \begin{bmatrix} \bar{x}_{\cdot 1} & \bar{x}_{\cdot 2} & \cdots & \bar{x}_{\cdot k} \\ \bar{x}_{\cdot 1} & \bar{x}_{\cdot 2} & \cdots & \bar{x}_{\cdot k} \\ \vdots & \vdots & & \vdots \\ \bar{x}_{\cdot 1} & \bar{x}_{\cdot 2} & \cdots & \bar{x}_{\cdot k} \end{bmatrix} \tag{10.14}$$

$$= \{x_{ij} - \bar{x}_{\cdot j}\} = \boldsymbol{X} - \mathbf{1}\bar{\boldsymbol{x}}'. \tag{10.15}$$

Inspection of (10.13) reveals that its jth column is

$$\begin{bmatrix} x_{1j} - \bar{x}_{\cdot j} \\ x_{2j} - \bar{x}_{\cdot j} \\ \vdots \\ x_{Nj} - \bar{x}_{\cdot j} \end{bmatrix} = \boldsymbol{x}_j - \frac{1}{N}\boldsymbol{J}_N\boldsymbol{x}_j = (\boldsymbol{I} - \bar{\boldsymbol{J}}_N)\boldsymbol{x}_j = \boldsymbol{C}_N\boldsymbol{x}_j, \tag{10.16}$$

where $\boldsymbol{C}_N$ is the centering matrix of Section 5.2:

$$\boldsymbol{C}_N = \boldsymbol{I} - \bar{\boldsymbol{J}}_N = \boldsymbol{I} - \frac{1}{N}\boldsymbol{J}_N = \boldsymbol{I} - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N{}'. \tag{10.17}$$

Hence (10.14) is

$$\{x_{ij} - \bar{x}_{\cdot j}\} = \boldsymbol{C}_N\boldsymbol{X}. \tag{10.18}$$

10.3.4 Corrected Sums of Squares and Products

From (10.16), using the symmetry and idempotency of $\boldsymbol{C}_N$, the sum of squares of observations in the jth column of $\boldsymbol{X}$, corrected for the mean, is

$$\sum_{i=1}^{N}(x_{ij} - \bar{x}_{\cdot j})^2 = \boldsymbol{x}_j{}'\boldsymbol{C}_N\boldsymbol{x}_j, \tag{10.19}$$

just like (5.19) in Section 5.2. Similarly, the corrected sum of cross-products of observations on the jth and tth variables is

$$\sum_{i=1}^{N}(x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t}) = \boldsymbol{x}_j{}'\boldsymbol{C}_N\boldsymbol{x}_t. \tag{10.20}$$

Arraying in a matrix all values of (10.19) and (10.20), for $j, t = 1, \ldots, k$, gives the matrix of corrected sums of squares and cross-products:

$$\left\{\sum_{i=1}^{N}(x_{ij} - \bar{x}_{.j})(x_{it} - \bar{x}_{.t})\right\} = \boldsymbol{X}'\boldsymbol{C}_N\boldsymbol{X}. \tag{10.21}$$

This is, of course, exactly analogous to (10.11) with $\boldsymbol{C}_N\boldsymbol{X}$ of (10.18) in place of $\boldsymbol{X}$, and using the symmetry and idempotency of $\boldsymbol{C}_N$:

$$(\boldsymbol{C}_N\boldsymbol{X})'\boldsymbol{C}_N\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{C}_N^2\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{C}_N\boldsymbol{X}.$$

Hence denoting the matrix of corrected sums of squares and products in (10.21) by $\boldsymbol{S}$, we have

$$\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{C}_N\boldsymbol{X}. \tag{10.22}$$

The matrix $\boldsymbol{S}$ (10.22) arises in multivariate statistics, where it is called the *Wishart matrix* (see Section 10.4); it also occurs in multiple regression (see Chapter 11).

Example 10.2 *When $\boldsymbol{X}$ is a vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)'$, equation (10.19) is*

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \boldsymbol{x}'\boldsymbol{C}_N\boldsymbol{x}, \tag{10.23}$$

as in (5.19). Because $\boldsymbol{C}_N$ is symmetric, we know from the spectral decomposition theorem in Section 6.2 that $\boldsymbol{C}_N$ can be expressed as $\boldsymbol{C}_N = \boldsymbol{PDP}'$, where $\boldsymbol{D}$ is the diagonal matrix of eigenvalues of $\boldsymbol{C}_N$ and $\boldsymbol{P}$ is an orthogonal matrix whose columns are the corresponding eigenvectors of $\boldsymbol{C}_N$. Hence,

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \boldsymbol{x}'\boldsymbol{PDP}'\boldsymbol{x} = \boldsymbol{y}'\boldsymbol{D}\boldsymbol{y}, \tag{10.24}$$

for $\boldsymbol{y} = \boldsymbol{P}'\boldsymbol{x}$. But, $\boldsymbol{C}_N$ is idempotent with $N - 1$ eigenvalues of 1 and one eigenvalue of zero. Hence,

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_{N-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{bmatrix}$$

and so from (10.24)

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N-1} y_i^2. \tag{10.25}$$

The importance of (10.25) in relation to sums of squares having χ^2-distributions is seen in the next section.

10.4 THE MULTIVARIATE NORMAL DISTRIBUTION

The probability density and moment generating functions for a normally distributed random variable x having mean μ and variance σ^2 are, respectively,

$$f(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \quad \text{and} \quad m_x(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}. \tag{10.26}$$

We say that x is normally distributed and write this as x is $N(\mu, \sigma^2)$ or $x \sim N(\mu, \sigma^2)$.

The multivariate analogue of (10.26) is for a vector $\boldsymbol{x}$, of k random variables, with $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and $\text{var}(\boldsymbol{x}) = \boldsymbol{V}$ as in (10.1) and (10.3). When all k variables are normally distributed, the joint density and moment generating functions, respectively, are

$$f(\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{\frac{k}{2}}|\boldsymbol{V}|^{\frac{1}{2}}} \quad \text{and} \quad m_{\boldsymbol{x}}(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{V}\boldsymbol{t}}. \tag{10.27}$$

The analogy between (10.26) and (10.27) is clear; and, of course, when $k = 1$, the functions in (10.27) reduce to those in (10.26). Notice in (10.27) that $f(\boldsymbol{x})$ exists only if $\boldsymbol{V}^{-1}$ does, that the denominator of $f(\boldsymbol{x})$ has $|\boldsymbol{V}|$ as the analogue of σ^2 in (10.26). A summary statement for (10.27) is $\boldsymbol{x} \sim N(\mu, \ V)$, just as $x \sim N(\mu, \ \sigma^2)$ is for (10.26).

Suppose $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \ \boldsymbol{V})$. We know from (10.4) and (10.5) that $\boldsymbol{Kx}$ has mean $\boldsymbol{K\mu}$ and dispersion matrix $\boldsymbol{KVK}'$. Using the result that a moment generating function does, under wide regularity conditions, uniquely define a corresponding density function, we use the moment generating function of (10.27) to show that

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V}) \quad \text{implies} \quad \boldsymbol{Kx} \sim N(\boldsymbol{K\mu}, \boldsymbol{KVK}'). \tag{10.28}$$

The definition of $m_{\boldsymbol{x}}(\boldsymbol{t})$ shown in (10.27) is equivalent to

$$m_{\boldsymbol{x}}(\boldsymbol{t}) = E\, e^{\boldsymbol{t}'\boldsymbol{x}}. \tag{10.29}$$

Therefore the moment generating function of $\boldsymbol{Kx}$ is

$$m_{\boldsymbol{Kx}}(\boldsymbol{t}) = E\, e^{\boldsymbol{t}'\boldsymbol{Kx}} = E\, e^{(\boldsymbol{K}'\boldsymbol{t})'\boldsymbol{x}}. \tag{10.30}$$

The last expression in (10.30) is, by analogy with (10.29), $m_{\boldsymbol{x}}(\boldsymbol{K}'\boldsymbol{t})$. Thus,

$$m_{\boldsymbol{Kx}}(\boldsymbol{t}) = m_{\boldsymbol{x}}(\boldsymbol{K}'\boldsymbol{t}). \tag{10.31}$$

Hence, using $\boldsymbol{K}'\boldsymbol{t}$ in place of $\boldsymbol{t}$ in (10.27) for the right-hand side of (10.31) gives

$$m_{\boldsymbol{Kx}}(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{K}\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{t}'\boldsymbol{KVK}'\boldsymbol{t}}. \tag{10.32}$$

Comparing (10.32) with (10.27) leads to conclusion (10.28).

Example 10.3 (continuation from Example 10.2) *Suppose that $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Then for $\boldsymbol{y} = \boldsymbol{P}'\boldsymbol{x}$ used in (10.25), with $\boldsymbol{P}$ orthogonal, (10.28) gives $\boldsymbol{y} \sim N(\boldsymbol{P}'\boldsymbol{0}, \boldsymbol{P}'\boldsymbol{IP})$, that is, $\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{I})$. This is always a consequence of an orthogonal transformation, that the distribution of an orthogonal transform of $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I})$ has the same distribution as $\boldsymbol{x}$.*

In the case of (10.25) there is a further consequence. The y_i's are normally distributed with zero mean and unit variance, so each y_i^2 is distributed as a χ^2 on 1 degree of freedom. The y_i's are also independent, so

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N-1} y_i^2$$

of (10.25) is distributed as the sum of $N - 1$ independent 1-degree-of-freedom χ^2's, that is, as χ^2 on $N - 1$ degrees of freedom.

The brevity of matrix notation is well illustrated in the preceding expressions and is further evident in considering marginal and conditional distributions of the multivariate normal distribution. Suppose $\boldsymbol{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{V}$ are partitioned conformably as

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix} \tag{10.33}$$

with $\boldsymbol{V}_{21} = (\boldsymbol{V}_{12})'$. Define $\boldsymbol{W}_{11} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}$, the Schur complement (see Section 4.12) of $\boldsymbol{V}_{22}$ in $\boldsymbol{V}$. Then it can be shown [*e.g., Searle* (*1971*), *Section 2.4f*] *that the marginal distribution of $\boldsymbol{x}_1$ is $\boldsymbol{x}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{V}_{11})$; and the conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$ is $N[\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{W}_{11}]$.*

These and numerous other uses of matrix algebra are widely found wherever multivariate statistical methods are discussed. For example, when the k random variables underlying the data matrix $\boldsymbol{X}_{N\times k}$ of (10.9) are normally distributed, $N(\boldsymbol{\mu}, \boldsymbol{V})$, then the $\frac{1}{2}k(k+1)$ distinct elements s_{ij} for $i \leq j = 1, \ldots, k$ of the Wishart matrix $\boldsymbol{S}$ of (10.22) are jointly distributed with probability density function

$$f(s_{11}, s_{12}, \ldots, s_{kk}) = \frac{|\boldsymbol{S}|^{\frac{1}{2}(N-k-1)} e^{-\frac{1}{2}\,tr(\boldsymbol{V}^{-1}\boldsymbol{S})}}{2^{\frac{1}{2}Nk}\pi^{\frac{1}{4}k(k-1)}\prod_{i=1}^{k}\Gamma\frac{1}{2}(N+1-i)}.$$

This is known as the Wishart density function.

10.5 QUADRATIC FORMS AND χ^2-DISTRIBUTIONS

Any sum of squares in an analysis of variance is a quadratic function of observations. Representing the observations by $\boldsymbol{x}$, a sum of squares is therefore the quadratic form $\boldsymbol{x}'\boldsymbol{Ax}$ for some symmetric matrix $\boldsymbol{A}$.

Hypothesis testing and variance component estimation often utilize expected values of sums of squares, that is, of $E(\mathbf{x}'\mathbf{A}\mathbf{x})$. When $\mathbf{x}$ has mean $\boldsymbol{\mu}$ and dispersion matrix $\mathbf{V}$ as in (10.1) and (10.3), we obtain $E(\mathbf{x}'\mathbf{A}\mathbf{x})$ by first writing $\mathbf{V} = E(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})' = E(\mathbf{x}\mathbf{x}') - \boldsymbol{\mu}\boldsymbol{\mu}'$, which is the multivariate analogue of the well-known univariate expression $\sigma^2 = E(x^2) - \mu^2$. Hence, $E(\mathbf{x}\mathbf{x}') = \mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}'$. Then, because $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a scalar, it equals its trace, $\mathbf{x}'\mathbf{A}\mathbf{x} = tr(\mathbf{x}'\mathbf{A}\mathbf{x})$, so that on using $tr(\mathbf{AB}) = tr(\mathbf{BA})$,

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = E\,tr(\mathbf{x}'\mathbf{A}\mathbf{x}) = E\,tr(\mathbf{A}\mathbf{x}\mathbf{x}') = tr[\mathbf{A}E(\mathbf{x}\mathbf{x}')],$$

thus giving

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = tr[\mathbf{A}(\mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}')] = tr(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}. \tag{10.34}$$

This result holds whether or not $\mathbf{x}$ is normally distributed. When it is, there is a further result for the variance of $\mathbf{x}'\mathbf{A}\mathbf{x}$:

$$\text{var}(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2tr(\mathbf{A}\mathbf{V})^2 + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}. \tag{10.35}$$

Results (10.34) and (10.35), the latter under normality, apply quite generally for any quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$. They are therefore available for providing the mean and variance of any sum of squares (expressed as a quadratic form) that may be of interest. Furthermore, generalizations to bilinear forms are also available, by writing a bilinear form as a quadratic form,

$$\mathbf{x}'\mathbf{A}\mathbf{y} = \frac{1}{2}\begin{bmatrix}\mathbf{x}' & \mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x} \\ \mathbf{y}\end{bmatrix} \tag{10.36}$$

and applying results like (10.34) and (10.35) to the right-hand side of (10.36). A variety of such results is given in Searle (1971, Sections 2.5 and 2.6).

F-statistics are ratios of mean squares or, more generally, of quadratic forms. And in order for such ratios to have an F-distribution, the numerator and denominator quadratic forms must be distributed independently of each other and each must have a χ^2 distribution.

The following sections provide key theorems concerning the distribution and independence of quadratic forms. Proofs of some of these theorems are quite lengthy. Details of these proofs will not be given here, but can be found in Khuri (2010, Sections 5.2– 5.5).

10.5.1 Distribution of Quadratic Forms

Theorem 10.1 *Let* $\mathbf{A}$ *be a symmetric matrix of order* $n \times n$ *and* $\mathbf{x}$ *be normally distributed as* $N(\boldsymbol{\mu}, \mathbf{V})$. *Then,* $\mathbf{x}'\mathbf{A}\mathbf{x}$ *can be expressed as*

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{k} \gamma_i W_i, \tag{10.37}$$

where $\gamma_1, \gamma_2, \ldots, \gamma_k$ are the distinct nonzero eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}$ (or, equivalently, the matrix $\boldsymbol{A}\boldsymbol{V}$) with multiplicities $\nu_1, \nu_2, \ldots, \nu_k$, respectively, and the W_i's are mutually independent such that $W_i \sim \chi^2_{\nu_i}(\theta_i)$, where

$$\theta_i = \boldsymbol{\mu}'\boldsymbol{V}^{-1/2}\boldsymbol{P}_i\boldsymbol{P}_i'\boldsymbol{V}^{-1/2}\boldsymbol{\mu}, \tag{10.38}$$

and $\boldsymbol{P}_i$ is a matrix of order $n \times \nu_i$ whose columns are orthonormal eigenvectors of $\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}$ corresponding to γ_i $(i = 1, 2, \ldots, k)$.

Theorem 10.2 *Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$ and $\boldsymbol{x}$ be normally distributed as $N(\boldsymbol{\mu}, \boldsymbol{V})$. A necessary and sufficient condition for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ to have the noncentral chi-squared distribution $\chi^2_r(\theta)$, where $\theta = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$, is that $\boldsymbol{A}\boldsymbol{V}$ be idempotent of rank r.*

Corollary 10.1 *Let $\boldsymbol{A}$ and $\boldsymbol{x}$ be the same as in Theorem 10.2. A necessary and sufficient condition for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ to have the $\chi^2_r(\theta)$ distribution is that $tr(\boldsymbol{A}\boldsymbol{V}) = tr[(\boldsymbol{A}\boldsymbol{V})^2] = r$, and $r(\boldsymbol{A}\boldsymbol{V}) = r$.*

Example 10.4 *The distribution of $\sum_{i=1}^n (x_i - \bar{x})^2$ has already been considered for $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$. Here we consider it for $\boldsymbol{x} \sim N(\mu\boldsymbol{1}_n, \sigma^2\boldsymbol{I}_n)$, that is, each $x_i \sim N(\mu, \sigma^2)$, with the x_i's being independent. Then, using*

$$\sum_{i=1}^n (x_i - \bar{x})^2/\sigma^2 = \boldsymbol{x}'[(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)/\sigma^2]\boldsymbol{x}, \tag{10.39}$$

we can invoke Theorem 10.2 to show that (10.39) has a χ^2-distribution. With $\boldsymbol{A}$ and $\boldsymbol{V}$ of the theorem being $(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)/\sigma^2$ and $\sigma^2\boldsymbol{I}_n$, respectively, $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n$, which is idempotent of rank $n - 1$. Furthermore, the noncentrality parameter is zero since

$$\theta = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} = \frac{1}{\sigma^2}\mu^2\boldsymbol{1}_n'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{1}_n = 0.$$

Theorem 10.2 gives

$$\sum_{i=1}^n (x_i - \bar{x})^2/\sigma^2 \sim \chi^2_{n-1}.$$

10.5.2 Independence of Quadratic Forms

Theorem 10.3 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices of order $n \times n$, and let $\boldsymbol{x}$ be normally distributed as $N(\boldsymbol{\mu}, \boldsymbol{V})$. A necessary and sufficient condition for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ to be independent is that*

$$\boldsymbol{A}\boldsymbol{V}\boldsymbol{B} = \boldsymbol{0}. \tag{10.40}$$

Theorem 10.4 *Let $\boldsymbol{A}$ be symmetric matrix of order $n \times n$, $\boldsymbol{B}$ be matrix order $m \times n$, and let $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$. A necessary and sufficient condition for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{B}\boldsymbol{x}$ to be independent is that*

$$\boldsymbol{B}\boldsymbol{V}\boldsymbol{A} = \boldsymbol{0}. \tag{10.41}$$

10.5.3 Independence and Chi-Squaredness of Several Quadratic Forms

Theorem 10.5 *Let $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I}_n)$ and $\boldsymbol{A}_i$ be a symmetric matrix of order $n \times n$ and rank k_i $(i = 1, 2, \ldots, p)$ such that $\boldsymbol{I}_n = \sum_{i=1}^{p} \boldsymbol{A}_i$. Then, any one of the following three conditions implies the other two:*

(a) $n = \sum_{i=1}^{p} k_i$, *that is, the ranks of $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_p$ sum to the rank of $\boldsymbol{I}_n$.*
(b) $\frac{1}{\sigma^2}\boldsymbol{x}'\boldsymbol{A}_i\boldsymbol{x} \sim \chi^2_{k_i}(\theta_i)$, *where* $\theta_i = \frac{1}{\sigma^2}\boldsymbol{\mu}'\boldsymbol{A}_i\boldsymbol{\mu}$, $i = 1, 2, \ldots, p$.
(c) $\boldsymbol{x}'\boldsymbol{A}_1\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{A}_2\boldsymbol{x}, \ldots, \boldsymbol{x}'\boldsymbol{A}_p\boldsymbol{x}$ *are mutually independent.*

Example 10.5 *Consider again Example 10.4 with $x_1, x_2, \ldots, x_n$ being a sample of independent and identically distributed as $N(\mu, \sigma^2)$. Let $\bar{x}$ and s^2 be the sample mean and sample variance, respectively, where*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

It was established that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. Now, using Theorem 10.4 we can show that $\bar{x} = \frac{1}{n}\boldsymbol{1}'_n\boldsymbol{x}$ and $s^2 = \frac{1}{n-1}\boldsymbol{x}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{x}$ are independent. In this case, $\boldsymbol{A} = \frac{1}{n-1}(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)$, $\boldsymbol{B} = \frac{1}{n}\boldsymbol{1}'_n$, and $\boldsymbol{x} \sim N(\mu\boldsymbol{1}_n, \sigma^2\boldsymbol{I}_n)$. Thus $\boldsymbol{V} = \sigma^2\boldsymbol{I}_n$. Hence,

$$\boldsymbol{B}\boldsymbol{V}\boldsymbol{A} = \frac{1}{n(n-1)}\boldsymbol{1}'_n(\sigma^2\boldsymbol{I}_n)(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n) = \boldsymbol{0}.$$

It follows that $\bar{x}$ and s^2 are independent.

10.5.4 The Moment and Cumulant Generating Functions for a Quadratic Form

Let $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ be a quadratic form in $\boldsymbol{x}$, where $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$ and $\boldsymbol{x}$ is distributed as $N(\boldsymbol{\mu}, \boldsymbol{V})$. The *moment generating function* of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is given by

$$\begin{aligned}\phi(t) &= E[\exp(t\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})] \\ &= \int_{R^n} \exp(t\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x},\end{aligned}$$

where $f(\boldsymbol{x})$ is the density function of $\boldsymbol{x}$ as shown in formula (10.27). In Khuri (2010, Section 5.1) it is shown that

$$\phi(t) = \frac{\exp\left\{-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{V}^{-1/2}\left[\boldsymbol{I}_n - (\boldsymbol{I}_n - 2\,t\,\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^{-1}\right]\boldsymbol{V}^{-1/2}\boldsymbol{\mu}\right\}}{|\boldsymbol{I}_n - 2\,t\,\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}|^{1/2}}. \tag{10.42}$$

This function is defined for $|\,t\,| < t_0$, where t_0 is given by

$$t_0 = \frac{1}{2\,max_i|\lambda_i|},$$

and λ_i $(i = 1, 2, \ldots, n)$ is the ith eigenvalue of the matrix $\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}$, or equivalently, the matrix $\boldsymbol{A}\boldsymbol{V}$.

Taking the natural logarithm of $\phi(t)$ produces the *cumulant generating function* of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, denoted by $\psi(t)$, which is given by the formula (see Khuri, 2010 for the actual derivation in Section 5.1),

$$\psi(t) = -\frac{1}{2}\log|\boldsymbol{I}_n - 2\,t\,\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}| - \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{V}^{-1/2}\left[\boldsymbol{I}_n - (\boldsymbol{I}_n - 2\,t\,\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^{-1}\right]\boldsymbol{V}^{-1/2}\boldsymbol{\mu}. \tag{10.43}$$

The rth *cumulant* of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, denoted by $\kappa_r(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})$, is the coefficient of $\frac{t^r}{r!}$ $(r = 1, 2, \ldots)$ in McLaurin's series expansion of $\psi(t)$. It can be shown that (see Khuri, 2010, Theorem 5.1)

$$\kappa_r(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = 2^{r-1}\,(r-1)!\,tr[(\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^r] + 2^{r-1}\,r!\,\boldsymbol{\mu}'\boldsymbol{V}^{-1/2}(\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^r\boldsymbol{V}^{-1/2}\boldsymbol{\mu},$$
$$r = 1, 2, \ldots \tag{10.44}$$

Note that

$$tr[(\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^r] = tr[(\boldsymbol{A}\boldsymbol{V})^r],$$

and

$$\boldsymbol{\mu}'\boldsymbol{V}^{-1/2}(\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2})^r\boldsymbol{V}^{-1/2}\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{A}\boldsymbol{V})^{r-1}\boldsymbol{A}\boldsymbol{\mu}.$$

Making the proper substitution in (10.44) gives

$$\kappa_r(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = 2^{r-1}(r-1)!\left\{tr[(\boldsymbol{A}\boldsymbol{V})^r] + r\boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{V})^{r-1}\boldsymbol{A}\boldsymbol{\mu}\right\},\ r = 1, 2, \ldots \tag{10.45}$$

Corollary 10.2 *If $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, then the mean and variance of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ are*

$$E(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + tr(\boldsymbol{A}\boldsymbol{V})$$
$$var(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = 4\,\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}\boldsymbol{\mu} + 2\,tr[(\boldsymbol{A}\boldsymbol{V})^2].$$

This follows directly from the fact that the mean and variance of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ are, respectively, the first and second cumulants of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$. Recall that the expression for $E(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})$ is valid even if $\boldsymbol{x}$ is not normally distributed.

10.6 COMPUTING THE CUMULATIVE DISTRIBUTION FUNCTION OF A QUADRATIC FORM

Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$ and let $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$. The cumulative distribution function of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is

$$F(u) = P(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq u). \tag{10.46}$$

The function $F(u)$ can be evaluated for a given value of u. Alternatively, one may compute the value of u for a given probability value in (10.46). Such a value of u is called the *pth quantile* of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, denote by u_p.

To evaluate $F(u)$, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ should be first expressed as a linear combination of mutually independent chi-squared variates as in formula (10.37) of Theorem 10.1. Doing so, we get

$$F(u) = P\left(\sum_{i=1}^{k} \gamma_i W_i \leq u\right), \tag{10.47}$$

where $\gamma_1, \gamma_2, \ldots, \gamma_k$ are the distinct nonzero eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}$ (or, equivalently, the matrix $\boldsymbol{A}\boldsymbol{V}$) with multiplicities $\nu_1, \nu_2, \ldots, \nu_k$, and the W_i's are mutually independent $\chi^2_{\nu_i}(\theta_i)$'s. Khuri(2010, Section 5.6) reported that the value of $F(u)$ can be easily calculated using a computer algorithm given be Davies (1980), which is based on a method proposed by Davies (1973). This algorithm is described in Davies (1980) as *Algorithm AS 155*, and can be easily accessed through *STATLIB*, which is an e-mail and file transfer protocol (FTP)-based retrieval system for statistical software.

Example 10.6 *Suppose that $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{V})$, where*

$$\boldsymbol{V} = \begin{bmatrix} 3 & 5 & 1 \\ 5 & 13 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{10.48}$$

Let $\boldsymbol{A}$ be the symmetric matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}.$$

The eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{A}\boldsymbol{V}^{1/2}$ (or $\boldsymbol{A}\boldsymbol{V}$) are −2.5473, 0.0338, 46.5135. We then have

$$F(u) = P(-2.5473\, W_1 + 0.0338\, W_2 + 46.5135\, W_3 \leq u),$$

where W_1, W_2, and W_3 are mutually independent variates, each distributed as χ^2_1. Using Davies' (1980) algorithm, quantiles of $-2.5473\, W_1 + 0.0338\, W_2 + 46.5135\, W_3$ can be obtained for given values of p. Some of these quantiles are presented in Table 10.1 (see also Table 5.3 in Khuri, 2010).

TABLE 10.1 Quantiles of $-2.5473\,W_1 + 0.0338\,W_2 + 46.5135\,W_3$

p	pth Quantile (u_p)
0.25	2.680 (first quartile)
0.50	18.835 (median)
0.75	59.155 (third quartile)
0.90	123.490 (90th percentile)
0.95	176.198 (95th percentile)

10.6.1 Ratios of Quadratic Forms

Ratios of quadratic forms are frequently used in *analysis of variance* (ANOVA) in form of ratios of mean squares obtained from an ANOVA table. Under certain conditions, such ratios serve as test statistics for testing hypotheses concerning certain parameters of an associated linear model. Consider, for example, the ratio

$$h(\boldsymbol{x}) = \frac{\boldsymbol{x}'\boldsymbol{A}_1\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}_2\boldsymbol{x}}, \tag{10.49}$$

where $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, and $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are symmetric matrices with $\boldsymbol{A}_2$ assumed to be positive semidefinite. The exact distribution of $h(\boldsymbol{x})$ is known only in some special cases, but is otherwise mathematically intractable. Several methods of approximating this distribution can be attempted in general, but the quality of approximation depends on the design used to generate the response data in a given experimental situation. It also depends on the assumptions made regarding the fitted model and the distribution of the data.

Let $G(u)$ denote the cumulative distribution function of $h(\boldsymbol{x})$. Then,

$$G(u) = P\left(\frac{\boldsymbol{x}'\boldsymbol{A}_1\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}_2\boldsymbol{x}} \leq u\right). \tag{10.50}$$

This can be written as

$$G(u) = P(\boldsymbol{x}'\boldsymbol{A}_u\boldsymbol{x} \leq 0), \tag{10.51}$$

where $\boldsymbol{A}_u = \boldsymbol{A}_1 - u\boldsymbol{A}_2$. Davies' algorithm referred to earlier in Section 10.6 can now be applied to obtain a tabulation of the values $G(u)$ for given values of u. For this purpose, we can express $\boldsymbol{x}'\boldsymbol{A}_u\boldsymbol{x}$ as a linear combination of mutually independent chi-squared variates, as in (10.47). Doing so, we get

$$G(u) = P\left(\sum_{i=1}^{l} \gamma_{ui}\, W_{ui} \leq 0\right), \tag{10.52}$$

where $\gamma_{u1}, \gamma_{u2}, \ldots, \gamma_{ul}$ are the distinct nonzero eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{A}_u\boldsymbol{V}^{1/2}$ (or, equivalently, $\boldsymbol{A}_u\boldsymbol{V}$) with multiplicities $\nu_{u1}, \nu_{u2}, \ldots, \nu_{ul}$, and the W_{ui}'s are mutually independent $\chi^2_{\nu_{ui}}(\theta_{ui})$'s.

TABLE 10.2 Values of $G(u)$ Using Formula (10.52)

u	$G(u)$
1.0	0.0
1.25	0.01516
1.50	0.01848
2.0	0.69573
2.75	0.99414
3.0	1.0

Example 10.7 *Consider the following ratio of quadratic forms:*

$$h(\boldsymbol{x}) = \frac{x_1^2 + 2\,x_2^2 + 3\,x_3^2}{x_1^2 + x_2^2 + x_3^2},$$

where $\boldsymbol{x} = (x_1, x_2, x_3)' \sim N(\boldsymbol{0}, \boldsymbol{V})$ and $\boldsymbol{V}$ is the same as in (10.48). In this case, $\boldsymbol{A}_1 = diag(1, 2, 3)$ and $\boldsymbol{A}_2 = \boldsymbol{I}_3$. Hence, $\boldsymbol{A}_u = \boldsymbol{A}_1 - u\boldsymbol{A}_2 = diag(1-u, 2-u, 3-u)$. For a given value of u, the eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{A}_u\boldsymbol{V}^{1/2}$ are obtained and used in formula (10.51). Davies' algorithm can now be used to obtain a tabulation of the values of $G(u)$. Some of these values are presented in Table 10.2.

REFERENCES

Davies, R. B. (1973). Numerical inversion of a characteristic function. *Biometrika*, 60, 415–417.

Davies, R. B. (1980). The distribution of a linear combination of χ^2 random variables. *Appl. Stat.*, 29, 323–333.

Khuri, A. I. (2010). *Linear Model Methodology*. Chapman & Hall/CRC, Boca Raton, FL.

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

EXERCISES

10.1 If the dispersion matrix of $\boldsymbol{x}$ is $\boldsymbol{I}$, and if the rows of $\boldsymbol{T}$ are orthonormal, prove that $\boldsymbol{y} = \boldsymbol{Tx}$ also has an identity matrix as its dispersion matrix.

10.2 Suppose θ is subtracted from every element of a data matrix. Show that the Wishart matrix is unchanged.

10.3 In a completely randomized experiment of n observations in each of a classes, let x_{ij} be the jth observation in the ith class for $i = 1, \ldots, a$ and $j = 1, \ldots, n$. Define $\boldsymbol{x}$ as the vector of observations in lexicon order

$$\boldsymbol{x}' = [x_{11}\ \ldots\ x_{1n}\ x_{21}\ \ldots\ x_{2n}\ \ldots \qquad\qquad \ldots\ x_{i1}\ \ldots x_{in}\ \ldots\ \ldots\ x_{a1}\ \ldots\ x_{an}]$$

and define the means

$$\bar{x}_{i.} = \sum_{j=1}^{n} \frac{x_{ij}}{n} \quad \text{and} \quad \bar{x}_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} \frac{x_{ij}}{an}.$$

Then in the analysis of variance of such data, the two sums of squares are

$$SS_A = n\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{x}_{..})^2 \quad \text{and} \quad SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})^2.$$

Show that

(a) $SS_A = \mathbf{x}'[\frac{1}{n}\mathbf{I}_a \otimes \mathbf{J}_n - \frac{1}{an}\mathbf{J}_a \otimes \mathbf{J}_n]\mathbf{x}$.

(b) $SS_E = \mathbf{x}'[\mathbf{I}_a \otimes \mathbf{I}_n - \frac{1}{n}\mathbf{I}_a \otimes \mathbf{J}_n]\mathbf{x}$.

(c) SS_A/σ^2 and SS_E/σ^2 are distributed independently.

(d) SS_A/σ^2 and SS_E/σ^2 are each distributed as a χ^2- distribution with $a-1$ and $a(n-1)$ degrees of freedom, respectively.

10.4 For $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ of order $n \times n$, show that for $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

(a) $E(\mathbf{y}'\mathbf{M}\mathbf{y}) = (n-r)\sigma^2$, where r is the rank of $\mathbf{X}$.

(b) $\mathbf{y}'\mathbf{M}\mathbf{y}$ and $\mathbf{y}'(\mathbf{I}-\mathbf{M})\mathbf{y}$ are independent.

(c) $\mathbf{y}'\mathbf{M}\mathbf{y}/\sigma^2 \sim \chi^2_{n-r}$.

(d) $\mathbf{y}'(\mathbf{I}-\mathbf{M})\mathbf{y}/\sigma^2 \sim \chi^2_r$ when $\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$.

10.5 Suppose

$$\text{var}[X\ Y\ Z]' = \begin{bmatrix} 8 & -7 & 2 \\ -7 & 10 & -3 \\ 2 & -3 & 1 \end{bmatrix},$$

where X, Y, and Z are three random variables.

(a) Prove that the 3×3 matrix is legitimate for this purpose.

(b) Show that $X + 2Y + 3Z$ and $Y + 4Z$ have variances 5 and 2, respectively. Are they correlated?

10.6 Prove that the determinant of $\mathbf{R}$ in (10.8) does not exceed unity, given that the variance-covariance matrix $\mathbf{V}$ is positive definite.

10.7 Why is $\mathbf{S}$ of (10.22) non-negative definite? Assuming it is positive definite, define an estimated correlation matrix $\hat{\mathbf{R}}$, analogous to (10.8), and show that its determinant does not exceed unity.

10.8 Let $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ and let n be the number of its elements. Show that the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ has the chi-squared distribution with n degrees of freedom if and only if $\mathbf{A} = \mathbf{V}^{-1}$.

10.9 Suppose that $\boldsymbol{x} = (x_1, x_2, x_3)'$ has the multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{V})$ such that

$$\begin{aligned} Q &= (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \\ &= 2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 - 2x_1x_3 - 4x_2x_3 - 6x_1 - 6x_2 + 10x_3 + 8 \end{aligned}$$

(a) Find $\boldsymbol{\mu}$ and $\boldsymbol{V}$.

(b) Find the moment generating function of $\boldsymbol{x}$.

10.10 The moment generating function of $\boldsymbol{x} = (x_1, x_2, x_3)'$ is given by

$$\phi_{\boldsymbol{x}}(\boldsymbol{t}) = exp(t_1 - t_2 + 2t_3 + t_1^2 + \frac{1}{2}t_2^2 + 2t_3^2 - \frac{1}{2}t_1t_2 - t_1t_3).$$

Find the value of $\mathrm{P}(x_1 + x_2 > x_3)$.

10.11 Let $\boldsymbol{W} = (\boldsymbol{x}', \boldsymbol{y}')' \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ have m, n elements, respectively. Let $\boldsymbol{A}$ be a constant matrix of order $m \times n$.

(a) Find $E(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y})$.

(b) Find $\mathrm{var}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y})$.

10.12 Show that if $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, then

$$\mathrm{cov}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}) = 2tr(\boldsymbol{AVBV}) + 4\boldsymbol{\mu}'\boldsymbol{AVB}\boldsymbol{\mu},$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are constant matrices.

11

Matrix Algebra of Full-Rank Linear Models

This chapter along with Chapter 12 are devoted to regression and linear models. Both topics are ideally suited to being described in matrix terminology and gain considerable clarity therefrom. It is hoped that these chapters will spark in the reader an interest in the merits of using matrix algebra for describing them.

The present chapter considers full-rank linear models, alternatively known as *linear regression models*, which are used to describe an empirical relationship between a response variable y and a set of control (or input) variables denoted by $x_1, x_2, \dots, x_k$. Such a relationship is developed on the basis of an experimental investigation where values of y are observed, or measured, in response to the application of particular settings of the control variables. The response y is a random variable assumed to have a continuous distribution and $x_1, x_2, \dots, x_k$ are nonstochastic variables whose values are determined by the experimenter and are usually measured on a continuous scale. Since y depends on the x_i's it is also called the *dependent variable* and the x-variables are correspondingly referred to as *independent variables*, or often as *regressor variables*. For example, y is the yield of a chemical reaction which depends on the control variables x_1, the reaction temperature, and x_2, the reaction time.

In general, the true relationship between y and $x_1, x_2, \dots, x_k$ is unknown. It is assumed, however, that it can be approximated by a polynomial model of degree d (≥ 1) in the control variables. It is customary to start out an experimental investigation by using a low-degree polynomial, such as a first-degree model, due to its simple form and low cost, particularly when little is known about the response in the initial stages of experimentation. This model is of the form

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon, \tag{11.1}$$

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are fixed unknown parameters, and ϵ is a random experimental error associated with the measured response at a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$ in a region of interest $\mathfrak{R}$. Model (11.1) is called a multiple linear regression model. If only one control variable is used in the model, then it is called a simple linear regression model. It is assumed that ϵ in (11.1) has a zero mean. Hence, the mean response $\mu(\boldsymbol{x})$, or expected value $E(y)$ of y, is of the form

$$\mu(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i. \tag{11.2}$$

During the course of experimentation and as more information becomes available about y, a higher-degree polynomial model may be needed in order to have a better representation of the true mean response. Quite often, a second-degree polynomial model of the following form is used:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \epsilon, \tag{11.3}$$

where the β's are fixed unknown parameters. In general, a polynomial representation of the relationship between y and the x_i's can be expressed as

$$y = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\beta} + \epsilon, \tag{11.4}$$

where $\boldsymbol{f}'(\boldsymbol{x})$ is a $1 \times p$ vector function whose first element is equal to one and the remaining elements are polynomial functions of the x_i's up to degree d (≥ 1), and $\boldsymbol{\beta}$ is a vector of p unknown parameters. The assumption that ϵ has a zero mean is based on the belief that, except for ϵ, the right-hand side of (11.4) represents the true mean response $\mu(\boldsymbol{x})$ at the point $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$, that is,

$$\mu(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\beta}. \tag{11.5}$$

Of course, such a belief may or may not be valid as possibly other functions of $\boldsymbol{x}$, including nonlinear functions, may be absent from the model. In this case, the model is said to suffer from *lack of fit* (LOF). Additional polynomial terms in $\boldsymbol{x}$ may be added to remove LOF, provided that added information on the response y are available that can sustain the fitting of an expanded model. Model (11.5) is called a *linear model* where "linear" indicates that the elements of $\boldsymbol{\beta}$ appear linearly in the model. All linear regression models can represented in such a fashion.

11.1 ESTIMATION OF β BY THE METHOD OF LEAST SQUARES

Estimation of $\boldsymbol{\beta}$ in model (11.5) requires the availability of n ($n > p$) observations on the response y where each observation is obtained as a result of a running an experiment using particular settings of the control variables $x_1, x_2, \ldots, x_k$. Let $\boldsymbol{x}_u = (x_{u1}, x_{u2}, \ldots, x_{uk})'$ denote a vector consisting of the settings x_{ui} of the control variables used in the uth run

($u = 1, 2, \ldots, n$). The subscript i keeps track of the control variables used in such a run ($i = 1, 2, \ldots, k$). Correspondingly, model (11.4) can be written as

$$y_u = \boldsymbol{f}'(\boldsymbol{x}_u)\,\boldsymbol{\beta} + \epsilon_u, \quad u = 1, 2, \ldots, n, \tag{11.6}$$

where y_u is the observed response value at $\boldsymbol{x}_u$ and ϵ_u is a random experimental error associated with y_u ($u = 1, 2, \ldots, n$). In matrix form, the totality of all the y_u's and corresponding settings of the control variables can be represented, using model (11.6), as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{11.7}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{X}$ is an $n \times p$ matrix whose uth row is $\boldsymbol{f}'(\boldsymbol{x}_u)$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$. The model matrix $\boldsymbol{X}$ is assumed to be of rank p, that is, has full column rank. Model (11.7) is then described as being of full rank. Additionally, we assume that $\mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and that $y_1, y_2, \ldots, y_n$ are uncorrelated and have variances equal to σ^2. Hence, the expected value (or mean) of $\boldsymbol{y}$ in model (11.7) is equal to $\boldsymbol{X}\boldsymbol{\beta}$ and its variance-covariance matrix is $\mathrm{var}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the identity matrix of order $n \times n$.

Under the above assumptions, the parameter vector $\boldsymbol{\beta}$ in model (11.7) can be estimated by using the *method of ordinary least squares* (OLS). This is accomplished by minimizing the square of the Euclidean norm of $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ with respect to $\boldsymbol{\beta}$, that is, minimizing the expression,

$$\begin{aligned} S(\boldsymbol{\beta}) &= \| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \|^2 \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\,\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}. \end{aligned} \tag{11.8}$$

A necessary condition for $S(\boldsymbol{\beta})$ to have a minimum at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ is that $\dfrac{\partial S(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} = \boldsymbol{0}$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, that is,

$$\left[\frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{y}'\boldsymbol{y} - 2\,\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right)\right]_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}} = \boldsymbol{0}. \tag{11.9}$$

Applying formulas (9.11) and (9.17) in Sections 9.3.1 and 9.3.3, respectively, we get

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{X}'\boldsymbol{y}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) = 2\,\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

Using these two formulas in (11.9) we get,

$$-\,2\,\boldsymbol{X}'\boldsymbol{y} + 2\,\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}. \tag{11.10}$$

Since $\boldsymbol{X}$ is of full column rank equal to p, the $p \times p$ matrix $\boldsymbol{X}'\boldsymbol{X}$ is of rank p and is therefore nonsingular (see property (xii) in Section 4.15). From (11.10) we conclude

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{y}. \tag{11.11}$$

Note that from (11.10) the *Hessian matrix* of second-order partial derivatives of $S(\boldsymbol{\beta})$ with respect to the elements of $\boldsymbol{\beta}$ (see Section 9.5.7) is equal to

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{\partial}{\partial \boldsymbol{\beta}} S(\boldsymbol{\beta})\right] &= \frac{\partial}{\partial \boldsymbol{\beta}'}[-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}] \\ &= 2\boldsymbol{X}'\boldsymbol{X},\end{aligned}$$

which is positive definite [see Theorem 5.3 (a)]. Hence, by Corollary 7.7.1 in Khuri (2003), $S(\boldsymbol{\beta})$ has a local minimum at $\hat{\boldsymbol{\beta}}$. Since it is the only solution to (11.10), $S(\boldsymbol{\beta})$ must achieve its absolute minimum at $\hat{\boldsymbol{\beta}}$. Alternatively, $S(\boldsymbol{\beta})$ can be shown to have an absolute minimum at $\hat{\boldsymbol{\beta}}$ by writing $S(\boldsymbol{\beta})$ in (11.8) as (see Exercise 11.7)

$$S(\boldsymbol{\beta}) = \| \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \|^2 + \| \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta} \|^2 . \tag{11.12}$$

We conclude that for all $\boldsymbol{\beta}$ in the parameter space,

$$\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \|^2 \geq \| \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \|^2 .$$

Equality is achieved if and only if $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. It follows that the absolute minimum of $S(\boldsymbol{\beta})$ is

$$\begin{aligned}S(\hat{\boldsymbol{\beta}}) &= \| \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \|^2 \\ &= \boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}.\end{aligned} \tag{11.13}$$

The right-hand side of (11.13) is called the *error (or residual) sum of squares*, and is denoted by SS_E. We thus have

$$SS_E = \boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}, \tag{11.14}$$

which has $n - p$ degrees of freedom since $\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is idempotent of rank $n - p$.

Example 11.1 *Suppose we have the following five sets of observations:*

i	y_i	x_{i1}	x_{i2}
1	62	2	6
2	60	9	10
3	57	6	4
4	48	3	13
5	23	5	2
$n = 5$	$y_{.} = 250$	$x_{.1} = 25$	$x_{.2} = 35$
	$\bar{y}_{.} = 50$	$\bar{x}_{.1} = 5$	$\bar{x}_{.2} = 7$

Then the estimated $\boldsymbol{\beta}$ is calculated as $\hat{\boldsymbol{\beta}}$ from (11.11) using

$$\boldsymbol{y} = \begin{bmatrix} 62 \\ 60 \\ 57 \\ 48 \\ 23 \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 2 & 6 \\ 1 & 9 & 10 \\ 1 & 6 & 4 \\ 1 & 3 & 13 \\ 1 & 5 & 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix}, \tag{11.15}$$

with

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \begin{bmatrix} 5 & 25 & 35 \\ 25 & 155 & 175 \\ 35 & 175 & 325 \end{bmatrix}^{-1} = \frac{1}{480}\begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix}. \tag{11.16}$$

Hence, from (11.11)

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \frac{1}{480}\begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix}\begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix} = \begin{bmatrix} 37 \\ \frac{1}{2} \\ \frac{3}{2} \end{bmatrix}. \tag{11.17}$$

11.1.1 Estimating the Mean Response and the Prediction Equation

Having obtained $\hat{\boldsymbol{\beta}}$, an estimate of the mean response in (11.5) is given by

$$\hat{\mu}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\beta}}.$$

This is also called the *predicted response* at $\boldsymbol{x}$, which is denoted by $\hat{y}(\boldsymbol{x})$ and the equation

$$\begin{aligned} \hat{y}(\boldsymbol{x}) &= \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \end{aligned} \tag{11.18}$$

is called the *prediction equation*. The $n \times 1$ vector $\hat{\boldsymbol{y}}$ consisting of the predicted response values at $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, where $\boldsymbol{x}_u$ is the setting of $\boldsymbol{x}$ at the uth experimental run ($u = 1, 2, \ldots, n$), is called the vector of predicted responses,

$$\begin{aligned} \hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \end{aligned} \tag{11.19}$$

Furthermore, the vector

$$\boldsymbol{y} - \hat{\boldsymbol{y}} = [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \tag{11.20}$$

is called the *residual vector*, which is orthogonal to $\hat{\boldsymbol{y}}$ since

$$\begin{aligned} \hat{\boldsymbol{y}}'(\boldsymbol{y} - \hat{\boldsymbol{y}}) &= \boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \\ &= 0. \end{aligned} \tag{11.21}$$

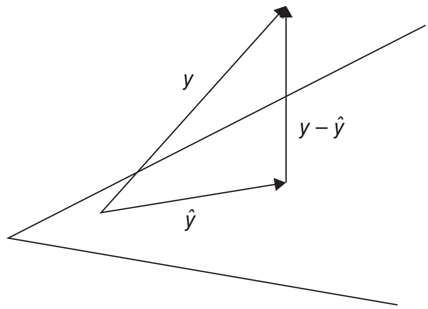


Figure 11.1 *Geometric Representation of Least Squares Estimation.*

It can be seen from (11.19) that $\hat{\boldsymbol{y}}$ belongs to the column space of $\boldsymbol{X}$ since it is a linear combination of the columns of $\boldsymbol{X}$. Furthermore, the mean response vector $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ also belongs to the same space and $\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0$, that is, residual vector $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to all vectors in the column space of $\boldsymbol{X}$. The idempotent matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ in (11.19) has therefore the effect of projecting the response vector $\boldsymbol{y}$ on the column space of $\boldsymbol{X}$ to produce $\hat{\boldsymbol{y}}$. This is demonstrated in Figure 11.1. The difference between $\boldsymbol{y}'\boldsymbol{y}$ and the error sum of squares SS_E in (11.14) is written as

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{y} - SS_E &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \end{aligned} \tag{11.22}$$

This is called the *regression sum of squares* and is denoted by SS_{reg}. Since the idempotent matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is of rank p, then SS_{reg} has p degrees of freedom. Hence, the total sum of squares, namely $\boldsymbol{y}'\boldsymbol{y}$, which is denoted by SS_T, is partitioned as

$$SS_T = SS_{reg} + SS_E. \tag{11.23}$$

Such a partitioning is usually displayed in an *analysis of variance* (ANOVA) table of the form:

TABLE 11.1 An ANOVA Table for a Regression Model

Source	DF (Degrees of Freedom)	SS	MS(Mean Square)
Regression	p	SS_{reg}	$MS_{reg} = SS_{reg}/p$
Error	$n-p$	SS_E	$MS_E = SS_E/(n-p)$
Total	n	SS_T	

Note that the mean square (MS) in Table 11.1 is obtained by dividing the sum of squares of a source by its degrees of freedom. It therefore represents the amount of variation per degree of freedom that is attributed to the source.

11.1.2 Partitioning of Total Variation Corrected for the Mean

Quite often in regression problems, one is interested in determining how much of the total variation in the response values, denoted by SS_{Tot}, is attributed to the model. The quantity $\sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the average of the response values, can be used to measure SS_{Tot}. The deviation of y_i from $\bar{y}$ can be partitioned as

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i), \quad i = 1, 2, \ldots, n. \tag{11.24}$$

Squaring both sides of (11.24) and taking the sum, we get

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{11.25}$$

The second term on the right-hand side of (11.25) is equal to zero since

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = \hat{\boldsymbol{y}}'(\boldsymbol{y} - \hat{\boldsymbol{y}}) - \bar{y}\mathbf{1}_n'(\boldsymbol{y} - \hat{\boldsymbol{y}}).$$

Both terms on the right-hand side are equal to zero since the residual vector $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to $\hat{\boldsymbol{y}}$, by virtue of (11.21), as well as to $\mathbf{1}_n$ because $\mathbf{1}_n$, being the first column of $\boldsymbol{X}$, belongs to the column space of $\boldsymbol{X}$. We then have

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{11.26}$$

Thus, SS_{Tot}, the total amount of variation in the response values, which is also referred to as the the sum of squares of the y_i's corrected for the mean, is partitioned into two components: $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, which measures the amount of variation in the response values that is explained by, or attributed to, the model, and $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, which is the error (or residual) sum of squares SS_E. The latter value represents variation due to the error term, that is, not accounted for by the model. The first component is called the *regression sum of squares corrected for the mean* and is denoted by SS_{Reg}. For the sake of simplicity, we shall just refer to it as the regression sum of squares. We thus have

$$SS_{Tot} = SS_{Reg} + SS_E. \tag{11.27}$$

Note that $SS_{Tot} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = SS_T - n\bar{y}^2$, and

$$SS_{Reg} = \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y}\sum_{i=1}^{n} \hat{y}_i + n\bar{y}^2. \tag{11.28}$$

Note that SS_T is the sum of squares of the n observation. But, $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$ by the fact that

$$\sum_{i=1}^{n}(\hat{y}_i - y_i) = (\hat{\boldsymbol{y}} - \boldsymbol{y})'\mathbf{1}_n = 0.$$

Hence, (11.28) is written as

$$SS_{Reg} = SS_{reg} - n\bar{y}^2.$$

It can be seen that (11.27) can be derived from (11.23) by subtracting $n\bar{y}^2$ from both sides. In (11.23), uncorrected sums of squares are used with SS_T, and in (11.27), corrected sums of squares are used with SS_{Tot}. The error sum of squares, SS_E, is the same in the two cases. Note that $n\bar{y}^2$ is actually the regression sum of squares for a model that only contains the parameter β_0 (see model 11.1). This is true since $\boldsymbol{y}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{y} = \frac{1}{n}(\mathbf{1}'\boldsymbol{y})^2 = n\bar{y}^2$.

Notation: The regression sum of squares SS_{reg}, which is not corrected for the mean, as well as the corrected one SS_{Reg}, can be partitioned into portions that help in assessing the contribution of a single regressor variable, or several regressor variables, to the amount of variation in the response values. It is quite common to use $R(\beta_0, \beta_1, \ldots, \beta_k)$ to denote SS_{reg} for model (11.1). Thus, $n\bar{y}^2$ is written as $R(\beta_0)$ and the regression sum of squares SS_{Reg}, which is corrected for the mean, is written as $R(\beta_0, \beta_1, \ldots, \beta_k) - R(\beta_0)$. Symbolically, this is expressed as $R(\beta_1, \beta_2, \ldots, \beta_k|\beta_0)$, where the vertical line denotes "in the presence of," which signals the increase in the regression sum of squares which results from adding the regressor variables $x_1, x_2, \ldots, x_k$ to a model which contains only the parameter β_0. By the same token, $R(.|.)$ measures the increase in the regression sum of squares resulting from adding the regressor variables corresponding to the parameters shown before the vertical line to a model that contains β_0 and the regressor variables whose parameters appear after the vertical line. For example, $R(\beta_2|\beta_0, \beta_1)$ represents the increase in the regression sums of squares that results from adding the regressor x_2 to a model that only contains x_1 and the unknown parameter β_0. It follows that the $R(.|.)$ notation can be used to provide a partitioning of SS_{reg}, or SS_{Reg}, into sequential regression sum of squares. For example, we have from model (11.1)

$$\begin{aligned} SS_{Reg} &= R(\beta_1, \beta_2, \ldots, \beta_k|\beta_0) \\ &= R(\beta_1|\beta_0) + R(\beta_2|\beta_0, \beta_1) + \cdots + R(\beta_k|\beta_0, \beta_1, \ldots, \beta_{k-1}). \end{aligned} \tag{11.29}$$

11.2 STATISTICAL PROPERTIES OF THE LEAST-SQUARES ESTIMATOR

Consider model (11.7) where it is assumed that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_n$. Hence, $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $\text{var}(\boldsymbol{y}) = \sigma^2\boldsymbol{I}_n$.

Several statistical properties associated with the least-squares estimator $\hat{\boldsymbol{\beta}}$ given in (11.11) will be described here.

11.2.1 Unbiasedness and Variances

The expected value of $\hat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$, that is, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of the parameter vector $\boldsymbol{\beta}$. This is true since by formula (10.4),

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}. \end{aligned}$$

The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ can be easily derived using formula (10.5) as follows:

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\text{var}(\boldsymbol{y})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.\end{aligned}$$

11.2.2 Estimating the Error Variance

The error sum of squares SS_E, defined in (11.14), is a quadratic form in $\boldsymbol{y}$ with $n-p$ degrees of freedom. Its expected value can be derived using formula (10.34). Since the mean of $\boldsymbol{y}$ is $\boldsymbol{X\beta}$ and its variance-covariance matrix is $\sigma^2\boldsymbol{I}_n$, we get by applying (10.34),

$$\begin{aligned}E(SS_E) &= \boldsymbol{\beta}'\boldsymbol{X}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X\beta} + \sigma^2 tr[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\\ &= \sigma^2(n-p).\end{aligned}$$

It follows that

$$E[SS_E/(n-p)] = \sigma^2. \tag{11.30}$$

This indicates that the error mean square, namely $MS_E = \frac{SS_E}{n-p}$, is an unbiased estimator of the error variance. This estimator is denoted by $\hat{\sigma}^2$.

Similarly, since SS_{reg} has p degrees of freedom, we have $E(MS_{reg}) = \frac{1}{p}\,\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} + \sigma^2$, where

$$MS_{reg} = \frac{1}{p}\,SS_{reg}. \tag{11.31}$$

This result also follows from applying formula (10.34) to SS_{reg}:

$$\begin{aligned}E(MS_{reg}) &= \frac{1}{p}\,E(SS_{reg})\\ &= \frac{1}{p}\,\left\{\boldsymbol{\beta}'\boldsymbol{X}'[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X\beta} + \sigma^2\,tr[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\right\}\\ &= \frac{1}{p}\,(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} + p\,\sigma^2)\\ &= \frac{1}{p}\,\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} + \sigma^2.\end{aligned}$$

Example 11.2 *Using (11.15) and (11.17) we get*

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 2 & 6\\ 1 & 9 & 10\\ 1 & 6 & 4\\ 1 & 3 & 13\\ 1 & 5 & 2\end{bmatrix}\begin{bmatrix} 37\\ \frac{1}{2}\\ 1\frac{1}{2}\end{bmatrix} = \begin{bmatrix} 47\\ 56\frac{1}{2}\\ 46\\ 58\\ 42\frac{1}{2}\end{bmatrix}.$$

Therefore, with $\mathbf{y}$ *from (11.15)*

$$(\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 62 \\ 60 \\ 57 \\ 48 \\ 23 \end{bmatrix} - \begin{bmatrix} 47 \\ 56\frac{1}{2} \\ 46 \\ 58 \\ 42\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 15 \\ 3\frac{1}{2} \\ 11 \\ -10 \\ -19\frac{1}{2} \end{bmatrix}$$

Substituting in $SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ *we get*

$$SS_E = 15^2 + \left(3\tfrac{1}{2}\right)^2 + 11^2 + 10^2 + \left(19\tfrac{1}{2}\right)^2 = 838\tfrac{1}{2}. \tag{11.32}$$

An alternative expression for SS_E *is*

$$\begin{aligned} SS_E &= \mathbf{y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}. \end{aligned}$$

It requires $\mathbf{y}'\mathbf{y} = \sum_{i=1}^{n} y_i^2 = 13526$ *obtainable from (11.15). Then with* $\hat{\boldsymbol{\beta}}'$ *from (11.17) and* $\mathbf{X}'\mathbf{y}$ *from (11.15), the value of* SS_E *is*

$$SS_E = 13526 - \begin{bmatrix} 37 & \frac{1}{2} & 1\frac{1}{2} \end{bmatrix} \begin{bmatrix} 250 \\ 1265 \\ 1870 \end{bmatrix} = 13526 - 12687\tfrac{1}{2} = 838\tfrac{1}{2}, \textit{ as before.}$$

From all of these and the fact that $\hat{\sigma}^2 = MS_E$ *we get*

$$\hat{\sigma}^2 = \frac{838\frac{1}{2}}{5-3} = 419\tfrac{1}{4}. \tag{11.33}$$

Example 11.3 (continued from Example 11.2) *From (11.32),* $SS_E = 838\frac{1}{2}$*, and from the data themselves in (11.15),* $SS_T = \sum_{i=1}^{n} y_i^2 = 13526$ *and* $R(\beta_0) = n\bar{y}^2 = 5(50^2) = 12500$*. Therefore, the summaries concerning* SS_{reg}*,* $SS_{Reg} = SS_{reg} - R(\beta_0)$*,* SS_E*,* SS_T*, and* $SS_{Tot} = SS_T - R(\beta_0)$ *are as shown in Table (11.2).*

TABLE 11.2 Partitioning of Sum of Squares

$SS_{reg} = 12687\frac{1}{2}$	$SS_{Reg} = 187\frac{1}{2}$
$SS_E = 838\frac{1}{2}$	$SS_E = 838\frac{1}{2}$
$SS_T = 13526$	$SS_{Tot} = 1026$

11.3 MULTIPLE CORRELATION COEFFICIENT

A measure of the goodness of fit of the regression is the multiple correlation coefficient, estimated as the product moment correlation between the observed y_i's and the predicted $\hat{y}_i$'s. Denoted by R. Its square, which is referred to as the *coefficient of determination*, is calculated as

$$R^2 = \frac{[\sum (y_i - \bar{y})(\hat{y}_\imath - \bar{\hat{y}})]^2}{\sum (y_\imath - \bar{y})^2 \sum (\hat{y}_\imath - \bar{\hat{y}})^2}.$$

Fortunately there is an easier formula, derived as follows.

First, in the denominator of R^2

$$\sum (y_\imath - \bar{y})^2 = \mathbf{y}'\mathbf{C}\mathbf{y} = \mathbf{y}'(\mathbf{I} - \bar{\mathbf{J}})\mathbf{y} = SS_{Tot},$$

where $\bar{\boldsymbol{J}} = \frac{1}{n}\boldsymbol{J}_n$ and $\boldsymbol{C} = \boldsymbol{I}_n - \bar{\boldsymbol{J}}$ is a centering matrix. Second, the numerator of R^2 is the square of

$$\begin{aligned}
(\mathbf{C}\mathbf{y})'\,\mathbf{C}\mathbf{y} &= \mathbf{y}'\mathbf{C}\mathbf{H}\mathbf{y} = \mathbf{y}'\left(\mathbf{I} - \bar{\mathbf{J}}\right)\mathbf{H}\mathbf{y} \\
&= \mathbf{y}'\left(\mathbf{H} - \bar{\mathbf{J}}\right)\mathbf{y} \qquad (11.34) \\
&= \mathbf{y}'\left[\left(\mathbf{I} - \bar{\mathbf{J}}\right) - (\mathbf{I} - \mathbf{H})\right]\mathbf{y} \\
&= SS_{Tot} - SS_E \\
&= SS_{Reg},
\end{aligned}$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. The expression in (11.34) is true because $\mathbf{1}_n$ is a column of $\boldsymbol{X}$, and so, since $\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{X}'$, $\mathbf{1}_n'\boldsymbol{H} = \mathbf{1}_n'$ and hence $\bar{\mathbf{J}}\boldsymbol{H} = \bar{\mathbf{J}} = \mathbf{H}\bar{\mathbf{J}}$. Finally, in the denominator of R^2,

$$\sum \left(\hat{y}_i - \bar{\hat{y}}\right)^2 = \hat{\mathbf{y}}'C\hat{\mathbf{y}} = \mathbf{y}'\mathbf{H}\mathbf{C}\mathbf{H}\mathbf{y} = \mathbf{y}'\left(\mathbf{H} - \bar{\mathbf{J}}\right)\mathbf{y} = SS_{Reg},$$

and so

$$R^2 = \frac{SS^2_{Reg}}{SS_{Tot}\,SS_{Reg}} = \frac{SS_{Reg}}{SS_{Tot}}. \qquad (11.35)$$

In the example,

$$R^2 = 187\tfrac{1}{2}/1026 = 0.183.$$

Intuitively the ratio SS_{Reg}/SS_{Tot} of (11.35) has appeal since it represents that fraction of the total (corrected) sum of squares which is accounted for by fitting the model—in this case fitting the regression. Thus although R has traditionally been thought of and used as a multiple correlation coefficient in some sense, its more frequent use nowadays is in the form (11.35), where it represents the fraction of the total sum of squares accounted for by fitting the model.

Example 11.4 *An experiment was conducted to study the relationship between three control variables, x_1, x_2, x_3, representing concentrations of three chemicals, and the yield y of some chemical compound. The data set is given in Table 11.3 where coded settings of the control variables are used.*

TABLE 11.3 Yield Data

x_1	x_2	x_3	y
−1.030	−1.602	−0.898	12.5988
0.915	0.488	−0.859	12.0838
0.871	−1.266	0.892	16.368
−0.955	0.478	0.982	14.2369
−0.937	−1.242	−0.898	9.3261
0.751	0.499	−0.619	17.0199
0.834	−1.093	0.743	13.4345
−0.951	0.388	0.854	16.4967
1.955	−0.472	0.012	14.6438
−2.166	−0.413	−0.039	20.8634
−0.551	0.059	−0.528	11.0421
−0.451	1.379	0.183	21.2098
0.152	1.209	0.083	25.5612
0.101	1.778	−0.009	33.3891
1.453	−0.353	0.183	15.5141

A second-degree model of the form,

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\ & + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \epsilon, \end{aligned} \tag{11.36}$$

was fitted to the data set using PROC GLM in SAS. The predicted response based on the least-squares estimates of the parameters in model (11.36) is given by

$$\begin{aligned} \hat{y} = {} & 14.491 + 1.306 x_1 + 5.137 x_2 + 7.065 x_3 + 6.390 x_1 x_2 + 2.376 x_1 x_3 - 2.538 x_2 x_3 \\ & + 1.191 x_1^2 + 2.433 x_2^2 - 3.218 x_3^2. \end{aligned} \tag{11.37}$$

The corresponding ANOVA table is displayed as Table 11.4. We note that in this table the sum of squares for "Regression" is actually the regression sum of squares corrected for the mean SS_{Reg} which is equal to 529.646. It has $p-1$ degrees of freedom where p is the number of columns of the model matrix $\mathbf{X}$, or the number of parameters in model (11.36). An estimate of the error variance is given by the error mean square, $MS_E = 2.297$. Furthermore, the coefficient of determination is $R^2 = SS_{Reg}/SS_{Tot} = 529.646/541.132 = 0.979$, that is,

TABLE 11.4 ANOVA Table for the Yield Data

Source	DF	SS	MS
Regression	9	529.646	58.850
Error	5	11.487	2.297
Corrected total	14	541.133	

almost 98% of the total (corrected) sum of squares is accounted for by fitting the model in (11.36).

11.4 STATISTICAL PROPERTIES UNDER THE NORMALITY ASSUMPTION

Let us now assume that $\boldsymbol{\epsilon}$ in model (11.7) has the normal distribution $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$. This results in a number of added properties concerning $\hat{\boldsymbol{\beta}}$. In particular, we have

(a) $\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$.
This follows directly from applying formula (10.28) to $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, which is a linear function of $\boldsymbol{y}$, and $\boldsymbol{y}$ is distributed as $N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$.

(b) $\hat{\boldsymbol{\beta}}$ and MS_E are independent.
The proof of this result follows from applying Theorem 10.4 to the linear form, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, and the quadratic form, $MS_E = \frac{1}{n-p}\boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}$. In this case, from applying the condition of independence given by (10.41), we get

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I}_n)\left\{\frac{1}{n-p}[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\right\} = \mathbf{0}.$$

The next six results describe properties concerning the distributions of SS_E and SS_{reg}, and their independence, in addition to the distribution of the regression sum of squares corrected for the mean SS_{Reg} and its independence from SS_E.

(c) $\frac{1}{\sigma^2} SS_E \sim \chi^2_{n-p}$.
This follows from applying Theorem 10.2 to SS_E given in (11.14) and the fact that $\frac{1}{\sigma^2}[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'](\sigma^2\boldsymbol{I}_n)$ is idempotent of rank $n-p$. The corresponding noncentrality parameter is zero since

$$\boldsymbol{\beta}'\boldsymbol{X}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X}\boldsymbol{\beta} = 0.$$

(d) $\frac{1}{\sigma^2} SS_{reg} \sim \chi^2_p(\theta)$, where $\theta = \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$.
Again, by applying Theorem 10.2 to SS_{reg} given in (11.22) we get the desired result since $\frac{1}{\sigma^2}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'](\sigma^2\boldsymbol{I}_n)$ is idempotent of rank p, and the noncentrality parameter is

$$\begin{aligned}\theta &= \boldsymbol{\beta}'\boldsymbol{X}'\left[\frac{1}{\sigma^2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]\boldsymbol{X}\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.\end{aligned}$$

(e) SS_{reg} and SS_E are independent.
By applying Theorem 10.3 to SS_{reg} and SS_E and using condition (10.40), we get

$$[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'](\sigma^2\boldsymbol{I}_n)[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] = \mathbf{0}.$$

This indicates that SS_{reg} and SS_E are independent.

(f) $\frac{1}{\sigma^2} SS_{Reg} \sim \chi^2_{p-1}(\theta)$ where $\theta = \frac{1}{\sigma^2}\boldsymbol{\beta_1}'\boldsymbol{X}_1{}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{X}_1\boldsymbol{\beta_1}$. Here, $\boldsymbol{\beta_1}$ is the portion of $\boldsymbol{\beta}$ left after removing its first element β_0, and $\boldsymbol{X}_1$ is the portion of $\boldsymbol{X}$ left after removing its first column $\boldsymbol{1}_n$.

This results from applying Theorem 10.2 to SS_{Reg} given in (11.28) since $\frac{1}{\sigma^2}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \frac{1}{n}\boldsymbol{J}_n](\sigma^2\boldsymbol{I}_n)$ is idempotent of rank $p-1$. The assertion that it is idempotent is true because

$$\left[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \frac{1}{n}\boldsymbol{J}_n\right]^2 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' -$$

$$\frac{1}{n}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{J}_n - \frac{1}{n}\boldsymbol{J}_n\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \frac{1}{n}\boldsymbol{J}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \frac{1}{n}\boldsymbol{J}_n.$$

The right-hand of the second equation follows from the fact that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}$. Hence,

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{1}_n,$$

since $\boldsymbol{1}_n$ is the first column of $\boldsymbol{X}$. This implies that

$$\begin{aligned} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{J}_n &= \boldsymbol{J}_n, \text{ and,} \\ \boldsymbol{J}_n\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' &= \boldsymbol{J}_n. \end{aligned} \tag{11.38}$$

The noncentrality parameter is equal to

$$\begin{aligned} \theta &= \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\left[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \frac{1}{n}\boldsymbol{J}_n\right]\boldsymbol{X}\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\left[\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right]\boldsymbol{X}\boldsymbol{\beta}. \end{aligned} \tag{11.39}$$

But,

$$\begin{aligned} \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}\boldsymbol{\beta} &= \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)[\boldsymbol{1}_n : \boldsymbol{X}_1]\boldsymbol{\beta} \\ &= \left[\boldsymbol{0} : \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}_1\right]\boldsymbol{\beta} \\ &= \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}_1\boldsymbol{\beta_1}. \end{aligned}$$

We conclude that

$$\theta = \frac{1}{\sigma^2}\boldsymbol{\beta_1}'\boldsymbol{X}_1'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}_1\boldsymbol{\beta_1}. \tag{11.40}$$

Note that $\theta = 0$ if and only if $\boldsymbol{\beta_1} = \boldsymbol{0}$. This is true because $\theta = 0$ if and only if

$$\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}_1\boldsymbol{\beta_1} = \boldsymbol{0}. \tag{11.41}$$

The equality in (11.41) implies that $\boldsymbol{\beta_1} = \mathbf{0}$. If not, then we must have

$$X_1\boldsymbol{\beta_1} = \frac{1}{n}\mathbf{1}_n(\mathbf{1}_n'X_1\boldsymbol{\beta_1}).$$

This implies that $\mathbf{1}_n$ belongs to the column space of X_1, which is a contradiction since $X = [\mathbf{1}_n : X_1]$ is of full column rank. Hence, $\boldsymbol{\beta_1} = \mathbf{0}$.

(g) $E(MS_{Reg}) = \frac{1}{p-1}[\boldsymbol{\beta_1}'X_1'(I_n - \frac{1}{n}J_n)X_1\boldsymbol{\beta_1} + \sigma^2$.
This follows directly from using the same argument as in part (f).

(h) SS_{Reg} and SS_E are independent.
In this case,

$$\begin{aligned}\left[X(X'X)^{-1}X' - \frac{1}{n}J_n\right][I_n - X(X'X)^{-1}X'] &= X(X'X)^{-1}X' - X(X'X)^{-1}X' \\ &\quad -\frac{1}{n}J_n + \frac{1}{n}J_nX(X'X)^{-1}X' \\ &= -\frac{1}{n}J_n + \frac{1}{n}J_n, \text{ by using (11.38)}, \\ &= \mathbf{0}. \end{aligned} \tag{11.42}$$

Thus SS_{Reg} and SS_E are independent by applying Theorem 10.3.

11.5 ANALYSIS OF VARIANCE

Partitioning the total sum of squares as shown in Tables 11.1 and 11.4 is the basis for analysis of variance. But using the customary F-statistics emanating from analysis of variance tables demands normality assumptions, and up to this point no need has arisen for, nor has any assumption been made about, probability distributional properties of the observations y_i or more particularly of the residuals ϵ_i. We now assume that they are normally distributed. This leads to sums of squares in the partitioning of total sum of squares having χ^2-distributions, and thence to ratios of mean squares having F-distributions from which come familiar tests of hypotheses based on F-statistics.

The basic distributional assumption to be made is that the error terms in $\boldsymbol{\epsilon}$ of the model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ are normally distributed with zero means and variance-covariance matrix $\sigma^2 I$, that is,

$$\boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2 I\right).$$

Distribution properties as shown in parts (a) through (h) follow at once.

Application of parts (c), (f), and (h) of the previous section to the sums of squares SS_{Reg} and SS_E yields the analysis of variance of Table 11.5.

Hypotheses tested by the F-statistics are as follows:

(a) $F(SS_{Reg}) = \frac{SS_{Reg}}{(p-1)MS_E}$ tests $H_0 : \boldsymbol{\beta_1} = \mathbf{0}$. This is true since by part (f) in the previous section, $\frac{1}{\sigma^2}SS_{Reg} \sim \chi^2_{p-1}(\theta)$ and the noncentrality parameter $\theta = \frac{1}{\sigma^2}\boldsymbol{\beta_1}'X_1'(I_n - \frac{1}{n}J_n)X_1\boldsymbol{\beta_1}$ is equal to zero if and only if $\boldsymbol{\beta_1} = \mathbf{0}$. Hence, $F(SS_{Reg})$ has under H_0 the

TABLE 11.5 Analysis of Variance for Regression for Model (11.7)

Source	DF	SS	MS	F-statistic
Regression (corrected)	$p-1$	SS_{Reg}	MS_{Reg}	$\frac{MS_{Reg}}{MS_E}$
Residual (error)	$n-p$	SS_E	MS_E	
Total (corrected)	$n-1$	SS_{Tot}		

F-distribution with $p-1$ and $n-p$ degrees of freedom. The test is significant at the α level if $F(SS_{Reg}) > F_{\alpha,p-1,n-p}$, where $F_{\alpha,p-1,n-p}$ is the upper α-percentage point of the F-distribution with $p-1$ and $n-p$ degrees of freedom.

(b) Likewise, based on part (d) of the previous section, $F(SS_{reg}) = \frac{SS_{reg}}{p\,MS_E}$ tests the hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The test is significant at the α-level if $F(SS_{reg}) > F_{\alpha,p,n-p}$.

Note that the test given by $F(SS_{Reg})$ tests if the variables in the fitted model are providing a significant amount of variation in the response values more than can be attributed to the error term. It is therefore the F-value reported in Table 11.5.

Using the data set for Example 11.4 with the corresponding ANOVA table (Table 11.4), we now present the same table with the addition of the F-statistic and the corresponding p-value. The latter value is called the *actual level of significance* of the F-test statistic as it represents the probability of exceeding the F-value shown in the ANOVA table. It also represents the upper tail area of the F-distribution with 9 and 5 degrees of freedom. Such a value can be compared against a given level of significance, for example, $\alpha = 0.05$ with a smaller value resulting in significance. Thus the p-value represents the smallest level the F-test statistic is significant at.

11.6 THE GAUSS–MARKOV THEOREM

The least-squares estimator of the parameter vector in a linear model has an optimal property concerning its variance. It is described in the following theorem, named after the German mathematician Carl Friedrich Gauss and the Russian mathematician Andrey (Andrei) Markov.

Theorem 11.1 *Consider model (11.7) where it is assumed that the experimental error vector $\boldsymbol{\epsilon}$ has a mean $\mathbf{0}$ and a variance-covariance matrix $\sigma^2\boldsymbol{I}_n$. Let $\mathbf{c}'\boldsymbol{\beta}$ be a linear function of the parameter vector $\boldsymbol{\beta}$, where $\mathbf{c}$ is a given nonzero constant vector. Then, $\mathbf{c}'\hat{\boldsymbol{\beta}} = \mathbf{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ has the smallest variance among all linear unbiased estimators of $\mathbf{c}'\boldsymbol{\beta}$. In this case, $\mathbf{c}'\hat{\boldsymbol{\beta}}$ is said to be the best linear unbiased estimator (BLUE) of $\mathbf{c}'\boldsymbol{\beta}$.*

TABLE 11.6 ANOVA Table for the Yield Data with the *F*-Statistic

Source	DF	SS	MS	F-Value	p-Value
Regression	9	529.646	58.850	25.62	0.0012
Error	5	11.487	2.297		
Corrected total	14	541.133			

Proof. By its definition, $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{y}$. It is also unbiased for $\boldsymbol{c}'\boldsymbol{\beta}$ since by Section 11.2.1 and formula (10.4), $E(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) = \boldsymbol{c}'\boldsymbol{\beta}$. It remains to show that $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ has the smallest variance among all linear unbiased estimators of $\boldsymbol{c}'\boldsymbol{\beta}$.

Let $\boldsymbol{t}'\boldsymbol{y}$ be any linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$, that is, $E(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{c}'\boldsymbol{\beta}$. This implies that

$$\boldsymbol{t}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{c}'\boldsymbol{\beta}, \tag{11.43}$$

which is true for all $\boldsymbol{\beta}$ in R^p, the p-dimensional Euclidean space. It follows from (11.43) that

$$\boldsymbol{t}'\boldsymbol{X} = \boldsymbol{c}'. \tag{11.44}$$

Using formula (10.5), the variance of $\boldsymbol{t}'\boldsymbol{y}$ is

$$\text{var}(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{t}'\boldsymbol{t}\,\sigma^2. \tag{11.45}$$

Furthermore, by using Section 11.2.1 and formula (10.5) again, we get

$$\text{var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) = \sigma^2\,\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}. \tag{11.46}$$

Using (11.44) in (11.46), we obtain

$$\text{var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) = \sigma^2\,\boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{t}. \tag{11.47}$$

From (11.45) and (11.47) we conclude that

$$\begin{aligned}\text{var}(\boldsymbol{t}'\boldsymbol{y}) - \text{var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) &= \sigma^2\,\boldsymbol{t}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{t}\\ &\geq 0.\end{aligned} \tag{11.48}$$

The inequality in (11.48) follows from the fact that the matrix $\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is idempotent, hence positive semidefinite. This shows that

$$\text{var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) \leq var(\boldsymbol{t}'\boldsymbol{y}). \tag{11.49}$$

Equality in (11.49) is attained if and only if $\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \boldsymbol{t}'\boldsymbol{y}$. This follows from noting in (11.48) that $\text{var}(\boldsymbol{t}'\boldsymbol{y}) = \text{var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$ if and only if

$$[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{t} = \boldsymbol{0},$$

or, equivalently, if and only if

$$\boldsymbol{t}' = \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'. \tag{11.50}$$

Using (11.44) in (11.50) we conclude

$$\begin{aligned}\boldsymbol{t}'\boldsymbol{y} &= \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{c}'\hat{\boldsymbol{\beta}}.\end{aligned}$$

■

The Gauss–Markov theorem is also applicable to vector linear functions of $\boldsymbol{\beta}$. This is given by the following corollary:

Corollary 11.1 *Let* $\boldsymbol{C}$ *be a given matrix of order* $q \times p$ *and rank* q $(\leq p)$*. Then,* $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ *is the* best linear unbiased estimator *(BLUE) of* $\boldsymbol{C}\boldsymbol{\beta}$*. Here, by best, it is meant that the matrix,*

$$\boldsymbol{B} = var(\boldsymbol{T}\boldsymbol{y}) - var(\boldsymbol{C}\hat{\boldsymbol{\beta}})$$

is positive semidefinite, where $\boldsymbol{T}\boldsymbol{y}$ *is any vector linear function of* $\boldsymbol{y}$ *that is unbiased for* $\boldsymbol{C}\boldsymbol{\beta}$*.*

Proof. The unbiasedness of $\boldsymbol{T}\boldsymbol{y}$ for $\boldsymbol{C}\boldsymbol{\beta}$ implies that $\boldsymbol{T}\boldsymbol{X} = \boldsymbol{C}$. The $q \times n$ matrix $\boldsymbol{T}$ must therefore be of rank q. This is true because $\boldsymbol{T}$, being of order $q \times n$, has rank $r(\boldsymbol{T}) \leq q$. But, since $\boldsymbol{C}$ is of rank q, then, $q = r(\boldsymbol{C}) = r(\boldsymbol{T}\boldsymbol{X}) \leq r(\boldsymbol{T})$. Hence, $r(\boldsymbol{T}) = q$.

Let us now show that the matrix $\boldsymbol{B}$ is positive semidefinite, that is, to show

$$\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} \geq 0, \tag{11.51}$$

for all $\boldsymbol{t} \in R^q$ with $\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} = 0$ for some $\boldsymbol{t} \neq \boldsymbol{0}$. We have that

$$\begin{aligned} \boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} &= \mathrm{var}(\boldsymbol{t}'\boldsymbol{T}\boldsymbol{y}) - \mathrm{var}(\boldsymbol{t}'\boldsymbol{C}\hat{\boldsymbol{\beta}}) \\ &= \mathrm{var}(\boldsymbol{u}'\boldsymbol{y}) - \mathrm{var}(\boldsymbol{z}'\hat{\boldsymbol{\beta}}), \end{aligned} \tag{11.52}$$

where $\boldsymbol{u}' = \boldsymbol{t}'\boldsymbol{T}$ and $\boldsymbol{z}' = \boldsymbol{t}'\boldsymbol{C}$. Note that $\boldsymbol{u}'\boldsymbol{y}$ is a linear function of $\boldsymbol{y}$ and is unbiased for $\boldsymbol{z}'\boldsymbol{\beta}$. The latter assertion is true because

$$\begin{aligned} E(\boldsymbol{u}'\boldsymbol{y}) &= \boldsymbol{u}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{t}'\boldsymbol{T}\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{t}'\boldsymbol{C}\boldsymbol{\beta} \\ &= \boldsymbol{z}'\boldsymbol{\beta}. \end{aligned} \tag{11.53}$$

From Theorem 11.1 we conclude that $\mathrm{var}(\boldsymbol{u}'\boldsymbol{y}) \geq \mathrm{var}(\boldsymbol{z}'\hat{\boldsymbol{\beta}})$ for all $\boldsymbol{t} \in R^q$, which implies (11.51). It remains to show that $\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} = 0$ for some $\boldsymbol{t} \neq \boldsymbol{0}$.

Suppose that $\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} = 0$. Then, from (11.52) we have

$$\sigma^2[\boldsymbol{u}'\boldsymbol{u} - \boldsymbol{z}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}] = 0. \tag{11.54}$$

But, $\boldsymbol{u}' = \boldsymbol{t}'\boldsymbol{T}$, $\boldsymbol{z}' = \boldsymbol{t}'\boldsymbol{C} = \boldsymbol{t}'\boldsymbol{T}\boldsymbol{X}$. Hence, (11.54) can be written as

$$\sigma^2\boldsymbol{t}'\boldsymbol{T}[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{T}'\boldsymbol{t} = 0. \tag{11.55}$$

We conclude that there is some nonzero vector $\boldsymbol{v} = \boldsymbol{T}'\boldsymbol{t}$ for which the equality in (11.55) is true, since the matrix $\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is positive semidefinite. For such a vector, $\boldsymbol{t} = (\boldsymbol{T}\boldsymbol{T}')^{-1}\boldsymbol{T}\boldsymbol{v} \neq \boldsymbol{0}$, which implies that $\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t} = 0$ for some nonzero $\boldsymbol{t}$. ■

A special case of Corollary 11.1 when $\boldsymbol{C} = \boldsymbol{I}_p$ shows that $\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{\beta}$.

11.6.1 Generalized Least-Squares Estimation

Consider again the linear model given in (11.7), but with a more general setup concerning the variance-covariance matrix of the random error vector $\boldsymbol{\epsilon}$. We assume that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{V}$, where $\boldsymbol{V}$ is a known positive definite matrix. Estimation of $\boldsymbol{\beta}$ in this case can be easily reduced to the earlier situation discussed in Section 11.1. This can be accomplished by multiplying both sides of (11.7) on the left by $\boldsymbol{V}^{-1/2}$. Doing so, we get

$$\boldsymbol{y}_g = \boldsymbol{X}_g \boldsymbol{\beta} + \boldsymbol{\epsilon}_g, \tag{11.56}$$

where $\boldsymbol{y}_g = \boldsymbol{V}^{-1/2}\boldsymbol{y}$, $\boldsymbol{X}_g = \boldsymbol{V}^{-1/2}\boldsymbol{X}$, and $\boldsymbol{\epsilon}_g = \boldsymbol{V}^{-1/2}\boldsymbol{\epsilon}$. Since $E(\boldsymbol{\epsilon_g}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon_g}) = \boldsymbol{V}^{-1/2}(\sigma^2 \boldsymbol{V})\boldsymbol{V}^{-1/2} = \sigma^2 \boldsymbol{I}_n$, then by the argument used in Section 11.1, the least-squares estimator of $\boldsymbol{\beta}$ in model (11.56) is

$$\hat{\boldsymbol{\beta}}_g = (\boldsymbol{X}_g' \boldsymbol{X}_g)^{-1} \boldsymbol{X}_g' \boldsymbol{y}_g.$$

This is equivalent to

$$\hat{\boldsymbol{\beta}}_g = (\boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{y}. \tag{11.57}$$

The vector $\hat{\boldsymbol{\beta}}_g$ is called the *generalized least-squares estimator*(GLSE) of $\boldsymbol{\beta}$. It is easy to see that it is unbiased for $\boldsymbol{\beta}$ and its variance-covariance matrix is of the form

$$\text{var}(\hat{\boldsymbol{\beta}}_g) = \sigma^2 (\boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}. \tag{11.58}$$

Furthermore, the Gauss–Markov Theorem is applicable in this more general case. For example, if $\boldsymbol{c}'\boldsymbol{\beta}$ is a linear function of $\boldsymbol{\beta}$, then its BLUE is $\boldsymbol{c}'\hat{\boldsymbol{\beta}}_g$ by Theorem 11.1. Also, if $\boldsymbol{C\beta}$ is vector linear function of $\boldsymbol{\beta}$, where $\boldsymbol{C}$ is a known matrix of order $q \times p$ and rank q $(\leq p)$, then its BLUE is $\boldsymbol{C}\hat{\boldsymbol{\beta}}_g$ by Corollary 11.1. In particular, $\hat{\boldsymbol{\beta}}_g$ is the BLUE of $\boldsymbol{\beta}$.

11.7 TESTING LINEAR HYPOTHESES

Consider model (11.7) under the assumption that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$. Suppose that it is of interest to test the general linear hypothesis,

$$H_0 : \boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m}, \tag{11.59}$$

against the alternative hypothesis,

$$H_a : \boldsymbol{K}'\boldsymbol{\beta} \neq \boldsymbol{m}, \tag{11.60}$$

where $\boldsymbol{K}'$ is a known matrix of order $r \times p$ and rank r $(\leq p)$, and $\boldsymbol{m}$ is a known vector of r elements.

Under H_0,

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{m}, \sigma^2 \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}],$$

Hence, under H_0,

$$\frac{1}{\sigma^2}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}) \sim \chi^2_r.$$

Furthermore, according to Property (b) in Section 11.4, $\hat{\boldsymbol{\beta}}$ is independent of MS_E and $\frac{n-p}{\sigma^2}MS_E \sim \chi^2_{n-p}$ by Property (c) in Section 11.4. It follows that under H_0, the statistic,

$$F = \frac{(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})}{rMS_E}, \tag{11.61}$$

has the F-distribution with r and $n - p$ degrees of freedom. Under the alternative hypothesis H_a,

$$\begin{aligned} E\{(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})\} &= (\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{m}_a - \boldsymbol{m}) + \\ &\quad tr\{[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]\sigma^2\} \\ &= (\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{m}_a - \boldsymbol{m}) \\ &\quad + r\sigma^2, \end{aligned} \tag{11.62}$$

where $\boldsymbol{m}_a$ is an alternative value of $\boldsymbol{K}'\boldsymbol{\beta}$ under H_a ($\boldsymbol{m}_a \neq \boldsymbol{m}$). It can be seen that

$$(\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{m}_a - \boldsymbol{m}) > 0,$$

because $\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}$ is a positive definite matrix. Thus, a large value of the test statistic F in (11.61) is significant, that is, H_0 is rejected at the α- level if $F \geq F_{\alpha,r,n-p}$.

Note that under the alternative value, $\boldsymbol{A\beta} = \boldsymbol{m}_a$, the test statistic F in (11.61) has the noncentral F-distribution with r and $n - p$ degrees of freedom and a noncentrality parameter given by

$$\theta = \frac{1}{\sigma^2}(\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{m}_a - \boldsymbol{m}).$$

Hence, the power of the test is

$$\text{Power} = P[F > F_{\alpha,r,n-p} \,|\, F \sim F_{r,n-p}(\theta)].$$

A special case of the hypothesis in (11.59) is

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0,$$

In this case, the statistic F in (11.61) takes the form

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{p\,MS_E}, \tag{11.63}$$

which, under H_0, has the F-distribution with p and $n - p$ degrees of freedom.

Example 11.5 *Consider again the data set of Example 11.1. Suppose that we wanted to test the joint hypothesis*

$$H: \begin{cases} \beta_1 &= \frac{1}{2} \\ \beta_1 + 3\beta_2 &= 3 \end{cases}, \quad \textit{equivalent to} \quad H: \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix}. \tag{11.64}$$

In this case,

$$\boldsymbol{K}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \quad \textit{and} \quad \boldsymbol{m} = \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix}.$$

We have

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 37 \\ \frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

and with $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ *from (11.16),*

$$\begin{aligned} \left[\boldsymbol{K}'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{K}\right]^{-1} &= \left\{ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \frac{1}{480} \begin{bmatrix} 790 & -80 & -42 \\ -80 & 16 & 0 \\ -42 & 0 & 6 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 3 \end{bmatrix} \right\}^{-1} \\ &= 480 \begin{bmatrix} 16 & 16 \\ 16 & 70 \end{bmatrix}^{-1} = \frac{5}{9} \begin{bmatrix} 70 & -16 \\ -16 & 16 \end{bmatrix}. \end{aligned}$$

The error mean square is given by (11.33). By making the substitution in (11.63), we get the value of test statistic, $F = \frac{35.556}{2\times 419.25} = 0.042$*, a nonsignificant result even at the 25% level* $(F_{0.25,2,5} = 1.85)$

11.7.1 The Use of the Likelihood Ratio Principle in Hypothesis Testing

An alternative method for deriving the test statistic for the null hypothesis H_0 in (11.59) is based on the *likelihood ratio approach* [see, e.g., Casella and Berger (2002, Section 8.2.1)]. It is structured on using the *likelihood function*, which for a given response vector $\boldsymbol{y}$, is the same as the density function of $\boldsymbol{y}$ under the assumption that $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$. By definition, the *likelihood ratio test statistic*, λ, for testing H_0 is given by

$$\lambda = \frac{\max_{H_0} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y})}{\max_{\boldsymbol{\beta}, \sigma^2} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y})}, \tag{11.65}$$

where $\max_{\boldsymbol{\beta}, \sigma^2} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y})$ is the maximum of the likelihood function maximized over the entire parameter space of $\boldsymbol{\beta} \in R^p$ and σ^2 $(0 < \sigma^2 < \infty)$, and $\max_{H_0} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y})$ denotes the maximum value of the likelihood function maximized over a restricted parameter space of $\boldsymbol{\beta}$ defined by $\boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m}$. This likelihood function is the density function described in (10.27) with a mean $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ and a variance-covariance matrix $\sigma^2 \boldsymbol{I}_n$. Since $\lambda \leq 1$, small values of λ lead to the rejection of H_0.

It can be shown (see Exercise 7) that $\max_{\boldsymbol{\beta},\sigma^2} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y})$ is given by

$$\max_{\boldsymbol{\beta},\sigma^2} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y}) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} e^{-n/2}, \tag{11.66}$$

where

$$\tilde{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \tag{11.67}$$

and $\hat{\boldsymbol{\beta}}$ is the least-squares estimator of $\boldsymbol{\beta}$.

It can also be shown that

$$\max_{H_0} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{y}) = \left[\frac{2\pi}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)\right]^{-n/2} e^{-n/2}, \tag{11.68}$$

where $\hat{\boldsymbol{\beta}}_r$ is the restricted (restricted under the null hypothesis) least-squares estimator of $\boldsymbol{\beta}$ given by

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}). \tag{11.69}$$

Using formulas (11.66) and (11.68), the expression in (11.65) can be written as

$$\lambda = \left[\frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)}\right]^{n/2}. \tag{11.70}$$

But,

$$(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = SS_E.$$

In addition, the minimum value of the residual sum of squares over the parameter space constrained by the hypothesis in (11.59) is attained when $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_r$ and is equal to

$$\begin{aligned}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r) &= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r)'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r) \\ &= SS_E + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r)'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r).\end{aligned}$$

Thus, from (11.69) we have

$$\begin{aligned}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r) &= SS_E + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r)'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_r) \\ &= SS_E + (\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1} \\ &\quad \times (\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}).\end{aligned} \tag{11.71}$$

We therefore have,

$$\lambda = \left[\frac{SS_E}{SS_E + Q}\right]^{n/2}, \tag{11.72}$$

where

$$Q = (\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}).$$

Since a small value of λ leads to the rejection of H_0 in (11.59), a large value of $\frac{Q}{SS_E}$, or equivalently, of $\frac{Q}{r\,MS_E}$, where $MS_E = \frac{SS_E}{n-p}$, will have the same effect. It can be recalled that $\frac{Q}{r\,MS_E}$ is the test statistic F in (11.61) for testing H_0. It follows that the likelihood ratio principle leads to the same test statistic as the one based on the distributional properties of $\hat{\boldsymbol{\beta}}$ under the normality assumption stated earlier.

11.7.2 Confidence Regions and Confidence Intervals

Let us now consider setting up a confidence region on a linear function of $\boldsymbol{\beta}$, namely, $\boldsymbol{K}'\boldsymbol{\beta}$, where $\boldsymbol{K}'$ is the matrix used in (11.59). As before, it is assumed that $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. In this case, a $(1-\alpha)100\%$ confidence region on such a linear function is given by

$$(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{K}'\boldsymbol{\beta})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{K}'\boldsymbol{\beta}) \le r\,MS_E\,F_{\alpha,r,n-p}. \tag{11.73}$$

If $\boldsymbol{K}'\boldsymbol{\beta}$ is a scalar of the form $\boldsymbol{k}'\boldsymbol{\beta}$, where $\boldsymbol{k}$ is a vector of p elements, then the t-distribution can be used to derive a $(1-\alpha)100\%$ confidence interval on $\boldsymbol{k}'\boldsymbol{\beta}$ that is of the form

$$\boldsymbol{k}'\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p}[\boldsymbol{k}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{k}\,MS_E]^{1/2}. \tag{11.74}$$

This follows from the fact that

$$\frac{\boldsymbol{k}'\hat{\boldsymbol{\beta}} - \boldsymbol{k}'\boldsymbol{\beta}}{[\boldsymbol{k}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{k}\,MS_E]^{1/2}}$$

has the t-distribution with $n-p$ degrees of freedom.

In particular, a $(1-\alpha)100\%$ confidence interval on the individual element, β_i, in $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ $(i = 1, 2, \ldots, p)$ is given by

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2},n-p}(a_{ii}\,MS_E)^{1/2}, \tag{11.75}$$

where a_{ii} is the diagonal element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ corresponding to β_i $(i = 1, 2, \ldots, p)$.

Another particular application of the confidence interval in (11.74) is the one concerning a *confidence interval on the mean response* $\boldsymbol{f}'(\boldsymbol{x}_0)\boldsymbol{\beta}$ at a given data point $\boldsymbol{x}_0$ (see model 11.5). In this case we have,

$$\boldsymbol{f}'(\boldsymbol{x}_0)\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p}[\boldsymbol{f}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x}_0)\,MS_E]^{1/2}. \tag{11.76}$$

There are other intervals concerning predictions of the response y within the region of experimentation. These intervals are called *prediction intervals* on new response observations. More specifically, a prediction interval on a new response observation y_0 at $\boldsymbol{x} = \boldsymbol{x}_0$, where $\boldsymbol{x}_0$ is a point inside an experimental region R, is an interval in which an experimenter can expect y_0 to fall. Under the assumption that $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, the derivation of such an interval is based on the fact that $y_0 - \hat{y}(\boldsymbol{x}_0)$ has the normal distribution with a zero mean and a variance given by $\sigma^2[1 + \boldsymbol{f}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}'(\boldsymbol{x}_0)]$ since $\hat{y}(\boldsymbol{x}_0) = \boldsymbol{f}'(\boldsymbol{x}_0)\hat{\boldsymbol{\beta}}$ and $y(\boldsymbol{x}_0)$, being not used in obtaining $\hat{y}(\boldsymbol{x}_0)$, must be independent of $\hat{y}(\boldsymbol{x}_0)$. Furthermore, since both $y(\boldsymbol{x}_0)$ and $\hat{y}(\boldsymbol{x}_0)$ are independent of the error mean square MS_E, we must have

$$\frac{y(\boldsymbol{x}_0) - \hat{y}(\boldsymbol{x}_0)}{\left\{MS_E[1 + \boldsymbol{f}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x}_0)]\right\}^{1/2}} \sim t_{n-p}.$$

It follows that the $(1 - \alpha)100\%$ prediction interval on y_0 is given by

$$\hat{y}(\boldsymbol{x}_0) \pm t_{\frac{\alpha}{2}, n-p} \left\{MS_E[1 + \boldsymbol{f}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x}_0)]\right\}^{1/2}. \tag{11.77}$$

Thus, this interval contains y_0 with a probability $1 - \alpha$. It should be noted this interval provides bounds that are expected to contain a random variable, namely, y_0, whereas the confidence interval on the mean response provides bounds that contain a fixed quantity, namely, the mean response $\boldsymbol{f}'(\boldsymbol{x}_0)\boldsymbol{\beta}$. The former should be wider than the latter since it accommodates the variability of y_0.

Example 11.6 *The heights and weights of 10 individuals randomly selected from a certain population are given in Table 11.7.*

TABLE 11.7 Height and Weight Data

Individual	Height (inches)	Weight (pounds)
1	64.75	111.25
2	70.98	135.60
3	71.30	154.09
4	69.01	143.35
5	68.75	145.19
6	67.32	122.95
7	68.46	140.32
8	69.37	135.25
9	69.21	114.56
10	67.82	121.83

The model fitted is the simple linear regression model,

$$y = \beta_0 + \beta_1 x + \epsilon,$$

where x = height and y = weight. The least-squares estimates of β_0 and β_0 are -230.915 and 5.289. Hence, the fitted model is

$$\hat{y} = -230.915 + 5.289x.$$

The corresponding ANOVA table is

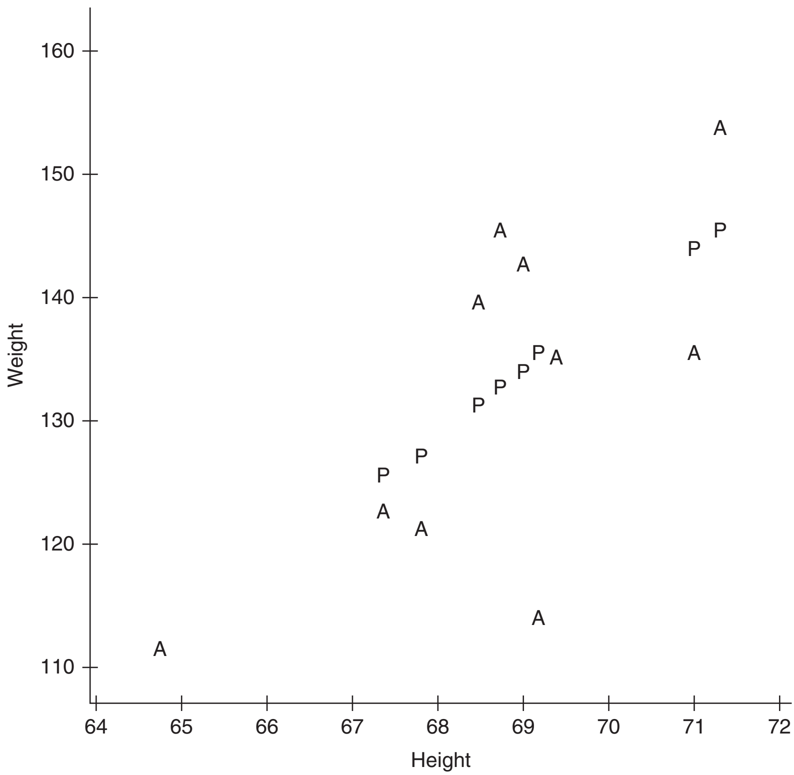
TABLE 11.8 The ANOVA Table

Source	DF	SS	MS	*F*-Value	*p*-Value
Model	1	870.188	870.188	7.47	0.0257
Error	8	931.444	116.431		
Corrected total	9	1801.632			

Note that the term "Model" is used by SAS to denote the regression corrected for the mean, and hence, the corresponding sum of squares in this ANOVA table is the same as SS_{Reg}, the regression sum of squares corrected for the mean. The corresponding F-value is significant at any level greater than or equal to 0.0257, which is the p-value. The null hypothesis tested by this F-value is $H_0 : \beta_1 = 0$. This indicates that there is a statistically significant linear trend between height and weight. The same hypothesis can also be tested by the corresponding t-statistic for the height parameter whose t-value is 2.733 with the same p-value, 0.0257. It should be noted that the square of this t-value is equal to the value of the F-statistic that corresponds to the regression sum of squares corrected for the mean.

The SAS system

Plot of weight*height. Symbol used is 'A'.
Plot of wtpred*height. Symbol used is 'P'.



Note: 2 obs hidden.

Figure 11.2 *Plot of Weight and Predicted Weight Values.*

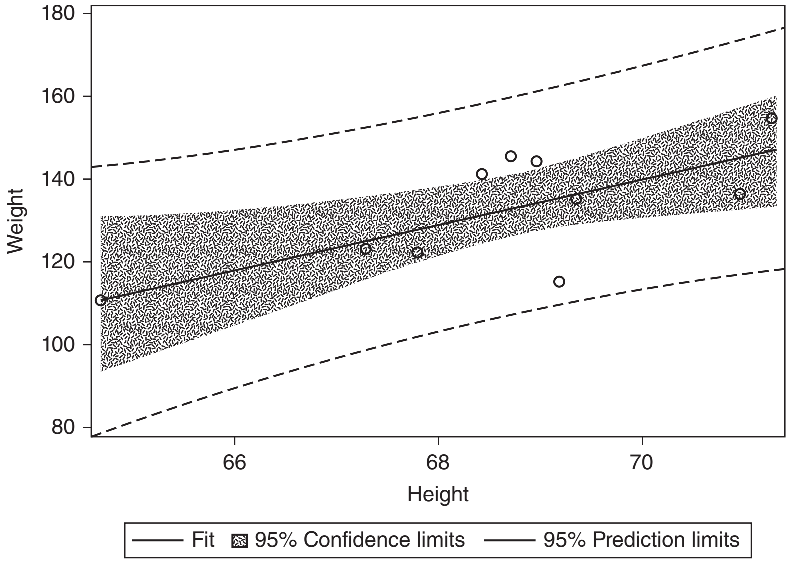


Figure 11.3 *Fit Plot for Weight with Confidence and Prediction Limits.*

A plot of the weight and predicted weight values is shown in Figure 11.2. In addition, the fit plot for weight versus height is given in Figure 11.3, which also shows the corresponding 95% confidence limits and the 95% prediction limits.

The SAS statements used to generate all the results in this example are

```
DATA ONE;
INPUT HEIGHT WEIGHT @@;
CARDS;
(include here the data following the order of the variables in the INPUT statement)
PROC PRINT;
PROC GLM;
MODEL WEIGHT = HEIGHT/p CLM;
OUTPUT OUT=NEW p=wtpred;
RUN;
PROC PLOT DATA = NEW;
PLOT WEIGHT*HEIGHT = 'A' wtpred*Height = 'P'/overlay;
RUN;
```

A detailed explanation of the above SAS statements will be given in Chapter 15 of Part III of this book.

Example 11.7 *Consider the data set of Example 11.4 with the corresponding ANOVA table 11.6. In the following table we have a listing of the parameters in model (11.36) along with their estimates, their t-statistic values, and p-values. Each p-value represents the probability of exceeding the absolute value of the corresponding t-statistic value. It can be seen that the parameters β_3, β_{12}, β_{23}, and β_{22} are significantly different from zero at a level less than 0.10.*

TABLE 11.9 Parameter Estimates and Corresponding t-Statistic Values

Parameter	Estimate	t-Statistic Value	p- Value
β_1	1.306	1.56	0.1794
β_2	5.137	1.67	0.1564
β_3	7.065	4.27	0.0079
β_{12}	6.390	3.89	0.0115
β_{13}	2.376	0.55	0.6087
β_{23}	−2.538	−2.71	0.0422
β_{11}	1.191	1.21	0.2809
β_{22}	2.433	2.14	0.0857
β_{33}	−3.218	−0.62	0.5644

In Table 11.10 we have a listing of the 95% confidence intervals on the mean response values at the 15 points of the yield data shown in Table 11.3

The corresponding 95% prediction intervals on new response observations at the 15 points of the yield data are given in Table 11.11.

The SAS statements used to obtain the results in Tables 11.9–11.11 are

DATA;
INPUT x_1 x_2 x_3 y;
CARDS;
(include here the data points in Table 11.3 of the yield data entered according to the order of variables in the INPUT statement, for example,
−1.030 −1.602 −0.898 12.5988
0.915 0.488 −0.859 12.0838
.
.
.

TABLE 11.10 95% Confidence Intervals on the Mean Response Values for the Yield Data

Data Point	Observed Resp.	Predicted Resp.	95% Conf. Limits for Mean Resp.	
1	12.590	12.576	9.082	16.069
2	12.084	13.376	10.080	16.672
3	16.368	15.336	12.016	18.656
4	14.237	14.840	11.641	18.039
5	9.326	9.352	6.258	12.447
6	17.020	15.781	13.137	18.425
7	13.435	14.882	12.026	17.738
8	16.497	15.244	12.467	18.021
9	14.644	13.973	10.472	17.474
10	20.863	21.141	17.266	25.015
11	11.042	10.380	7.106	13.654
12	21.210	22.230	19.013	25.446
13	25.561	25.998	23.090	28.906
14	33.389	32.582	29.391	35.774
15	15.514	16.097	13.257	18.938

TABLE 11.11 95% Prediction Intervals on New Response Values for the Yield Data

Data Point	Observed Resp.	Predicted Resp.	95% Prediction limits for New Resp.	
1	12.590	12.576	7.342	17.809
2	12.084	13.376	8.272	18.480
3	16.368	15.336	10.217	20.455
4	14.237	14.840	9.799	19.881
5	9.326	9.352	4.377	14.328
6	17.020	15.781	11.072	20.489
7	13.435	14.882	10.051	19.713
8	16.497	15.244	10.460	20.029
9	14.644	13.973	8.735	19.211
10	20.863	21.141	15.646	26.635
11	11.042	10.380	5.291	15.469
12	21.210	22.230	17.177	27.282
13	25.561	25.998	21.136	30.860
14	33.389	32.582	27.546	37.619
15	15.514	16.097	11.275	20.919

1.453 −0.353 0.183 15.5141)
PROC PRINT;
PROC GLM;
MODEL y =x_1 x_2 x_3 $x_1 * x_2$ $x_1 * x_3$ $x_2 * x_3$ x_1^2 $x_2^2 x_3^2$/p CLM;
PROC GLM;
MODEL y =x_1 x_2 x_3 $x_1 * x_2$ $x_1 * x_3$ $x_2 * x_3$ x_1^2 $x_2^2 x_3^2$/p CLI;
RUN;

In the first MODEL statement, CLM refers to "confidence limits for mean" response, and in the second MODEL statement, CLI refers to "prediction limits for an individual" response. The "p" option in both MODEL statements is needed in order to print observed response and predicted response values for each data point. The need to have two MODEL statements is due to the fact that CLI should not be used with CLM in the same MODEL statement; it is ignored if CLM is also specified. Uppercase letters used in writing the SAS statements are not necessary. They were used here just for convenience to provide a clearer representation of such statements. We note that each statement ends in a semicolon, but no semicolon is needed in the listing of the data. Note also that each model statement should be preceded by a PROC GLM statement.

11.8 FITTING SUBSETS OF THE *x*-VARIABLES

Consider a regression analysis of plant growth on five x-variables: x_1 and x_2 are climatic variables (e.g., mean daily temperature and rainfall) and x_3, x_4, and x_5 are seedling measurements at time of transplanting (e.g., age, leaf development, and height of transplant). We might wish to ascertain whether or not the two climatic variables make a significant contribution to the regression. If they do not, the regression would be based solely on the seedling measurements x_3, x_4, and x_5, and for just these three variables a regression could be estimated. In other words, we are interested in testing the hypothesis $H : \beta_1 = \beta_2 = 0$.

Suppose that instead of writing the regression as

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

we change the order of the terms to have

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_0 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

and with this order partition $\boldsymbol{X\beta}$ of $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ to be

$$E(\boldsymbol{y}) = \boldsymbol{X}_1\boldsymbol{\beta_1} + \boldsymbol{X}_2\boldsymbol{\beta_2} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix} \tag{11.78}$$

with

$$\boldsymbol{\beta_1}' = [\beta_1 \quad \beta_2] \quad \text{and} \quad \boldsymbol{\beta_2}' = [\beta_0 \quad \beta_3 \quad \beta_4 \quad \beta_5].$$

Then the hypothesis $H: \beta_1 = \beta_2 = 0$ is

$$H: \boldsymbol{\beta_1} = \boldsymbol{0}. \tag{11.79}$$

Hypotheses of this nature are frequently of interest: testing whether some subset of the k x-variables is contributing to the regression.

We consider the partitioned model (11.78) and hypothesis (11.79) in the general case where $\boldsymbol{\beta_1}$ is taken as having k_1 elements and $\boldsymbol{\beta_2}$ having k_2, with $k_1 + k_2 = k + 1$. Then for testing $H: \boldsymbol{\beta_1} = \boldsymbol{0}$ we have $\boldsymbol{K}' = [\boldsymbol{I} \quad \boldsymbol{0}]$, and $\boldsymbol{m} = \boldsymbol{0}$, so that $\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} = \hat{\boldsymbol{\beta}}_1$ and, on defining

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}^{-1} \equiv \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix} \quad \text{with} \quad \boldsymbol{T}_{21} = (\boldsymbol{T}_{12})', \tag{11.80}$$

$$\left[\boldsymbol{K}'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{K}\right]^{-1} = \left\{\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix}\right\}^{-1} = \boldsymbol{T}_{11}^{-1}. \tag{11.81}$$

Hence, the F-statistic for testing $H: \boldsymbol{\beta_1} = \boldsymbol{0}$ is

$$F(\boldsymbol{\beta_1}) = Q_1/(k_1 MS_E) \quad \text{with} \quad Q_1 = \hat{\boldsymbol{\beta}}_1{}'\boldsymbol{T}_{11}^{-1}\hat{\boldsymbol{\beta}}_1. \tag{11.82}$$

11.9 THE USE OF THE *R*(.|.) NOTATION IN HYPOTHESIS TESTING

Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta_1} + \boldsymbol{X}_2\boldsymbol{\beta_2} + \boldsymbol{\epsilon}, \tag{11.83}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are matrices of orders $n \times k_1$ and $n \times k_2$, respectively. The reduction in the sum squares due to fitting the full model in (11.83) after having fitted a reduced model that contains only $\boldsymbol{X}_2\boldsymbol{\beta}_2$ is written as $R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$. By definition,

$$R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2) = R(\boldsymbol{\beta}) - R(\boldsymbol{\beta}_2), \tag{11.84}$$

where $R(\boldsymbol{\beta})$ is the regression sum of squares for the full model in (11.83) and $R(\boldsymbol{\beta}_2)$ is the regression sum of squares due to fitting the reduced model which contains $\boldsymbol{X}_2\boldsymbol{\beta}_2$. We thus have

$$R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2) = \boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}, \tag{11.85}$$

where $\boldsymbol{X} = [\boldsymbol{X}_1 : \boldsymbol{X}_2]$.

Note that the difference between the two regression sums of squares in (11.85) is also equal to the difference between the residual sum of squares over the parameter space constrained by the hypothesis in (11.79), and the residual sum of squares for the full model. By the result given in (11.71), the latter difference is then equal to $(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})$, where here $\boldsymbol{K}' = [\boldsymbol{I}\ \ \boldsymbol{0}]$. But, this difference is the same as the numerator sum of squares of the F-test given in (11.61) for testing the null hypothesis $H : \boldsymbol{\beta}_1 = \boldsymbol{0}$, that is, the F-test statistic $F(\boldsymbol{\beta}_1)$ given in (11.82).

Using Property (h) in Section 4.12 concerning the inverse of the partitioned matrix $\boldsymbol{X}'\boldsymbol{X}$ shown in (11.80), we have

$$\begin{aligned}
\boldsymbol{T}_{11} &= [\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1]^{-1} \\
\boldsymbol{T}_{12} &= -\boldsymbol{T}_{11}\boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}.
\end{aligned}$$

Then, it can be verified that

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_1 &= \boldsymbol{T}_{11}\boldsymbol{X}_1'\boldsymbol{y} + \boldsymbol{T}_{12}\boldsymbol{X}_2'\boldsymbol{y} \\
&= (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y},
\end{aligned} \tag{11.86}$$

where $\boldsymbol{M}_2 = \boldsymbol{I} - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$. Substituting the expressions for $\hat{\boldsymbol{\beta}}_1$ and $\boldsymbol{T}_{11}$ in formula (11.82), we find that Q_1 in (11.82) reduces to

$$Q_1 = \boldsymbol{y}'\boldsymbol{M}_2\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y},$$

which is equal to $R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$.

Example 11.8 *Consider the sums of squares in Table 11.2. Then,* $SS_{reg} = R(\beta_0, \beta_1, \beta_2) =$ 12687.5. *For* $E(y) = \beta_0 + \beta_2 x_2$, *the reduction in sum of squares* $R(\beta_0.\beta_2)$ *is 12680.*

Therefore,

$$R(\beta_1|\beta_0, \beta_2) = R(\beta_0, \beta_1, \beta_2) - R(\beta_0, \beta_2) = 7.50,$$

which is the numerator of $F(\beta_1) = \frac{7.50}{MS_E} = \frac{7.50}{419.25} = 0.018.$

REFERENCES

Khuri, A. I. (2003). *Advanced Calculus with Applications in Statistics*, 2nd ed. John Wiley & Sons, New York.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd ed. Duxbury, Pacific Grove, CA.

EXERCISES

11.1 Consider the model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times p$ of rank p, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. Show that

$$E\left(\frac{1}{MS_E}\right) = \frac{n-p}{\sigma^2(n-p-2)},$$

provided that $n - p > 2$.

11.2 Consider the simple linear regression model,

$$y_u = \beta_0 + \beta x_u + \epsilon_u, \quad u = 1, 2, \ldots, n,$$

where $\epsilon_u \sim N(0, \sigma^2)$, $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent, and $n > 4$.

Show that $\frac{n-4}{(n-2)MS_E}$ is an unbiased estimator of $\frac{1}{\sigma^2}$.

11.3 Consider again Exercise 2. Show that $\text{var}(\hat{\beta}_0)$ is a minimum if $x_1, x_2, \ldots, x_n$ are chosen so that $\bar{x} = 0$, where $\bar{x} = \frac{1}{n}\sum_{u=1}^{n} x_u$.

11.4 Consider the following data set used to fit the regression model

$$y = \beta_0 + \beta_1 x_{u1} + \beta_2 x_{u2} + \epsilon_u, \quad u = 1, 2, \ldots, n,$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and distributed as $N(0, \sigma^2)$:

x_1	0.87	0.202	0.203	0.198	0.730	0.510	0.205	0.670	0.205	0.271	0.203	0.264
x_2	1.690	1.170	1.170	1.210	1.630	1.590	1.140	1.920	1.220	1.710	1.160	1.370
y	12.10	5.50	4.60	4.50	10.80	4.90	6.00	4.20	5.30	6.70	4.00	6.10

(a) Find the least-squares estimates of the parameters in the model as well as the error mean square, MS_E.

(b) Test the hypothesis $H_0 : \beta_0 - 2\beta_2 = 0, \beta_0 = 8.0$ using $\alpha = 0.05$. Find also the power of this test, given that $\sigma^2 = 1$ and $H_a : \beta_0 - 2\beta_2 = 1$ and $\beta_0 = 10$.

(c) Obtain individual confidence intervals on β_0, β_1, β_2 using a 95% confidence coefficient for each.

Note: For the benefit of the reader, the following SAS statements can be used to obtain the solutions to parts (a), (b), and (c):

```
DATA ONE:
INPUT X1 X2 YY;
```

```
CARDS;
(Include here the data set following the order of variables as in the INPUT
 statement).
PROC PRINT;
PROC GLM;
MODEL YY = X1 X2;
DATA TWO;
SET ONE;
RUN;
PROC IML;
USE TWO; READ ALL INTO X11;
ONE=J(12,1,1);
XX1=X11[1:12,{1 2}];
X = ONE||XX1;
PRINT X;
NN=X‘*X; N = INV(NN);
PRINT N;
Y = X11[1:12,3];
PRINT Y;
MSE = Y‘*(I(12) - X*N*X‘)*Y/9;
PRINT MSE;
BETA = N*X‘*Y;
K = {1 1,0 0,-2 0};
PRINT K;
M = {0,8};
MA = {1, 10};
F = (K‘*BETA - M)‘*(INV(K‘*N*K))*(K‘*BETA - M)/(2*MSE);
PRINT F;
LAMBDA = (MA - M)‘*(INV(K‘*N*K))*(MA - M);
PRINT LAMBDA;
POWER = 1 - PROBF(4.26,2,9,LAMBDA);
PRINT POWER;
```

A detailed explanation of the above SAS statements will be given in Chapter 15 of Part III of this book.

11.5 Consider the simple linear regression model, $y = \beta_0 + \beta_1 x + \epsilon$. Let $\mu = \beta_0 + \beta_1 x$ be the mean response at x. Let x_0 be the value of x at which $\mu = 0$. Show how to obtain a $(1-\alpha)100\%$ confidence set on x_0 (assume that the usual assumptions of normality, independence, and equality of error variances are satisfied).

11.6 Let $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V})$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\boldsymbol{X}$ is an $n \times p$ matrix of rank p $(n > p)$, and $\boldsymbol{V}$ is a known variance-covariance matrix.

What is the distribution of

(a) $\boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$?

(b) $\boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}$?

11.7 Show that $S(\boldsymbol{\beta})$ has an absolute minimum at $\hat{\boldsymbol{\beta}}$ by writing $S(\boldsymbol{\beta})$ in (11.8) as in (11.12).

11.8 Verify the equality given in (11.86).

11.9 Consider the model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon.$$

A 2^k factorial design was used to fit this model. The design settings were coded using the ± 1 convention. This design was augmented with n_0 center-point replications. Let $\bar{y}_1$ be the average of the response values at the points of the 2^k design, and let $\bar{y}_0$ denote the average of the replicated response values at the design center. The random errors are assumed to be independently distributed as normal variables with variances equal to σ^2.

(a) If the true mean response is represented by a second-degree model of the form

$$\mu = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2,$$

then show that

$$E(\bar{y}_1 - \bar{y}_0)^2 = \left(\frac{1}{2^k} + \frac{1}{n_0}\right)\sigma^2 + \left(\sum_{i=1}^{k} \beta_{ii}\right)^2.$$

(b) Show that if the model given in part (a) is the true mean response, then the error mean square, MS_E, from fitting the first-degree model, is no longer an unbiased estimator of σ^2. In this case, find an unbiased estimator of σ^2.

(c) Give a test statistic for testing the hypothesis

$$H_0 : \sum_{i=1}^{k} \beta_{ii} = 0 \text{ versus } H_a : \sum_{i=1}^{k} \beta_{ii} \neq 0.$$

11.10 Consider again Exercise 9. Let $\hat{y}(\boldsymbol{x}_1)$ and $\hat{y}(\boldsymbol{x}_2)$ denote the predicted response values at the points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ inside the experimental region using the first-degree model. Show that $\text{var}[\hat{y}(\boldsymbol{x}_1) - \hat{y}(\boldsymbol{x}_2)]$ depends on the distances of the points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ from the center and the angle θ subtended by these points at the center (that is, θ is the angle between the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$) (the design used is still 2^k plus n_0 center points).

11.11 Consider the linear model given in (11.7), where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$, the matrix $\boldsymbol{\Sigma}$ is known and assumed to be nonsingular. Let $MS_E = \boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)$. Find the expected value of MS_E and show that

$$E(MS_E) \leq \frac{tr(\boldsymbol{\Sigma})}{n-p},$$

where $tr(\boldsymbol{\Sigma})$ is the trace of $\boldsymbol{\Sigma}$.

11.12 Consider again Exercise 11. The random error vector $\boldsymbol{\epsilon}$ is assumed to have the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$. Let $\theta = \boldsymbol{\lambda}'\boldsymbol{\beta}$, where $\boldsymbol{\lambda}$ is a given constant vector. Obtain a $(1-\alpha)100\%$ confidence interval on θ.

11.13 Consider a linear model as in (11.7), where it is assumed that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$. Let $y(\boldsymbol{x}_0)$ be a "new" response value at a point $\boldsymbol{x}_0$ (that is, $y(\boldsymbol{x}_0)$ is not one of the response values used to fit this model). Let $\hat{y}(\boldsymbol{x}_0)$ denote the predicted response at the same point $\boldsymbol{x}_0$ (the prediction results from the least-squares fit of the model).

(i) Show how to compute the following probability for given $\boldsymbol{x}_0$ and $\boldsymbol{X}$:

$$P\left[\frac{\hat{y}(\boldsymbol{x}_0) - y(\boldsymbol{x}_0)}{\sqrt{MS_E}} > 1\right],$$

where MS_E is the error mean square.

(ii) Let $d = E[(\hat{y}(\boldsymbol{x}_0) - y(\boldsymbol{x}_0))^2]$.

(a) Express d in terms of σ^2, $\boldsymbol{x}_0$, and $\boldsymbol{X}$.

(b) Find an unbiased estimator $\hat{d}$ of d.

(c) Show how to compute the probability (for given $\boldsymbol{x}_0$ and $\boldsymbol{X}$), $P[\hat{d} > 2\sigma^2]$.

12

Less-Than-Full-Rank Linear Models

The general subject of linear statistical models that are less than full rank is a large one and is presented extensively in many texts such as Kempthorne (1952), Scheffé (1959), Graybill (1961, 1976), Searle (1971, 1987), Winer (1971), Rao (1973), Seber (1977), and more recently, Jorgensen (1993), Hinkelmann and Kempthorne (1994), Rao and Toutenburg (1995), Christensen (1996), Khuri et al. (1998), Sengupta and Jammalamadaka (2003), and Khuri (2010), to name but a few. It is a subject of great importance in data analysis, because of its close association with analysis of variance. It is also very amenable to presentation in matrix terminology which, in turn, brings considerable clarity to the subject. Nevertheless, because the topic is so vast just its fringe is touched here. Great reliance is placed on results given in Chapter 11, and considerable use is made of the generalized inverse of a matrix discussed in Chapter 8.

To those familiar with linear models, it can be said that the main objective of this chapter is to present a simple and general method of deriving estimable functions and their unique, least-squares (minimum variance, linear, unbiased) estimators. The method is based entirely on the properties of generalized inverses and is applicable to any situation whatever, whether the data are balanced or unbalanced.

12.1 GENERAL DESCRIPTION

To understand what is meant by a less-than-full linear model, we consider the linear model,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{12.1}$$

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

where the Matrix $\boldsymbol{X}$ does not have a full column rank. As before in Chapter 11, $\boldsymbol{X}$ is a known matrix of order $n \times p$, but its rank is r that is strictly less p. In this case, the matrix $\boldsymbol{X}'\boldsymbol{X}$, which is of order $p \times p$ and rank r, is singular. Consequently, it cannot be used to estimate $\boldsymbol{\beta}$.

Examples of models of this kind are numerous and include analysis of variance (ANOVA) models such as crossed or nested classification models. The simplest crossed classification model is the one-way model,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, k; \; j = 1, 2. \ldots, n_i. \tag{12.2}$$

It can be verified that the matrix $\boldsymbol{X}$ for this model is $[\mathbf{1}_{n_.} : \oplus_{i=1}^{k} \mathbf{1}_{n_i}]$, which is of order $n_. \times (k+1)$ and rank k $(n_. = \sum_{i=1}^{k} n_i)$, and the corresponding vector of unknown parameters is $\boldsymbol{\beta} = (\mu, \alpha_i, \alpha_2, \ldots, \alpha_k)'$.

A simple illustration serves to introduce the general idea of linear models and the matrix notation applicable to them.

Example 12.1 *Federer (1955) reports an analysis of rubber-producing plants called guayule, for which the plant weights were available for 54 plants of 3 different kinds, 27 of them normal, 15 off-types, and 12 aberrants. We will consider just 6 plants for purposes of illustration, 3 normals, 2 off-types, and 1 aberrant, as shown in Table (12.1).*

TABLE 12.1 Weights of Six Plants

	Normal	Off-Type	Aberrant	
	101	84	32	
	105	88		
	94			
Total	300	172	32	Grand total = 504

For the entries in this table, let y_{ij} denote the weight of the jth plant of the ith type, i taking values 1, 2, 3 for normal, off-type, and aberrant, respectively, and $j = 1, 2, \ldots, n_i$, where n_i is the number of observations in the ith type. The problem is to estimate the effect of type of plant on weight of plant. To do this we assume the observation y_{ij} is the sum of three parts

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \tag{12.3}$$

where μ represents the population mean of the weight of plant, α_i is the effect of type i on weight, and ϵ_{ij} is a random residual term peculiar to the observation y_{ij}.

As in regression, it is assumed that the ϵ_{ij}'s are independently distributed with zero mean, that is, $E(\epsilon_{ij}) = 0$. Then $E(y_{ij}) = \mu + \alpha_i$. It is also assumed that each ϵ_{ij} has the same variance, σ^2, so that the variance-covariance matrix of the vector of ϵ-terms is $\sigma^2 \boldsymbol{I}_{n_.}$. This is clearly a linear model since it is based on the assumption that y_{ij} is the simple sum of its three parts μ, α_i, and ϵ_{ij}.

The problem is to estimate μ and the α_i's, and also σ^2, the variance of the error terms. We will find that not all of the terms μ and α_i can be estimated satisfactorily, only certain linear functions of them can. At first thought this may seem to be a matter for concern;

sometimes it is, but on many occasions it is not, because the number of linear functions that can be estimated is large and frequently includes those in which we are interested; that is, differences between the effects, such as $\alpha_1 - \alpha_2$. Sometimes, however, functions of interest cannot be estimated, because of a paucity of data. On all occasions, though, a method is needed for ascertaining which functions can be estimated and which cannot. This is provided in what follows.

To develop the method of estimation, we write down the six observations in terms of equation (12.1) of the model:

$$\begin{aligned}
101 = y_{11} &= \mu + \alpha_1 \qquad\qquad\quad + \epsilon_{11}\\
105 = y_{12} &= \mu + \alpha_1 \qquad\qquad\quad + \epsilon_{12}\\
94 = y_{13} &= \mu + \alpha_1 \qquad\qquad\quad + \epsilon_{13}\\
84 = y_{21} &= \mu \qquad\quad + \alpha_2 \qquad + \epsilon_{21}\\
88 = y_{22} &= \mu \qquad\quad + \alpha_2 \qquad + \epsilon_{22}\\
32 = y_{31} &= \mu \qquad\qquad\quad + \alpha_3 + \epsilon_{31}.
\end{aligned}$$

These equations are easily written in matrix form as

$$\begin{bmatrix} 101\\105\\94\\84\\88\\32 \end{bmatrix} = \begin{bmatrix} y_{11}\\y_{12}\\y_{13}\\y_{21}\\y_{22}\\y_{31} \end{bmatrix} = \begin{bmatrix} 1&1&0&0\\1&1&0&0\\1&1&0&0\\1&0&1&0\\1&0&1&0\\1&0&0&1 \end{bmatrix} \begin{bmatrix} \mu\\ \alpha_1\\ \alpha_2\\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11}\\ \epsilon_{12}\\ \epsilon_{13}\\ \epsilon_{21}\\ \epsilon_{22}\\ \epsilon_{31} \end{bmatrix}, \tag{12.4}$$

which is precisely the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{12.5}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{X}$ is the matrix of 0's and 1's, $\boldsymbol{\beta}$ is the vector of parameters to be considered,

$$\boldsymbol{\beta} = (\mu,\ \alpha_1,\ \alpha_2,\ \alpha_3)',$$

and $\boldsymbol{\epsilon}$ is the vector of error terms

Note that (12.5) is exactly the same form as (11.7). In both cases, $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are vectors of observations and errors, respectively, and in both cases $\boldsymbol{\beta}$ is a vector of parameters whose estimates we seek. Also, the properties of $\boldsymbol{\epsilon}$ are the same: $E(\boldsymbol{\epsilon}) = \mathbf{0}$, so that $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{1}_n$. The only difference is in the form of $\mathbf{X}$: in regression it is a matrix of observations on x-variables, whereas in (12.5) its elements are either 0 or 1, depending on the absence or presence of particular terms of the model in each y_{ij} observation. But, with respect to applying the principle of least squares for estimating the elements of $\boldsymbol{\beta}$, there is no difference at all between (12.5) and (11.7). Hence we can go directly to (11.10) to obtain an equation for $\hat{\boldsymbol{\beta}}$, the least-squares estimator of $\boldsymbol{\beta}$:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{12.6}$$

In regression analysis the solution $\hat{\boldsymbol{\beta}}$ *was obtained from this as* $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ *where* $\mathbf{X}'\mathbf{X}$ *was nonsingular and therefore had an inverse. Now, however,* $\mathbf{X}'\mathbf{X}$ *has no inverse and we have to use the methods of generalized inverses to obtain a solution.*

12.2 THE NORMAL EQUATIONS

12.2.1 A General Form

Equations (12.6) are called the *normal equations*. Before discussing their solution let us look briefly at their form. First, the vector of parameters, $\boldsymbol{\beta}$: it is the vector of all the elements of the model, in this case the elements μ, α_1, α_2, and, α_3. And this is so in general; for example, if data can be arranged in rows and columns according to two different classifications, elements of the vector $\boldsymbol{\beta}$ can be μ and terms representing, say, a row effects and b column effects and up to ab interaction effects. Therefore $\boldsymbol{\beta}$ can have as many as $1 + a + b + ab$ elements in that case.

Second, the matrix $\mathbf{X}$: it is called the *incidence matrix* (or, for data from designed experiments, the *design matrix*), because the location of the 0's and 1's throughout its elements represents the incidence of terms of the model among the observations—and hence of the classifications in which the observations lie. This is particularly evident if one writes $\mathbf{X}$ as a two-way table with the parameters as headings to the columns and the observations as labels for the rows, as illustrated in Table 12.2.

In Table 12.2, as in equations (12.4), it is clear that the sum of the last three columns equals the first column. (This is so because every y_{ij} observation contains μ and so the first column of $\mathbf{X}$ is all 1's; and every y_{ij} also contains just one α and so the sum of the last three columns is also all 1's.) Thus $\mathbf{X}$ is not of full column rank.

Now consider the normal equations (12.6). It involves $\mathbf{X}'\mathbf{X}$, which is obviously square and symmetric. Its elements are the inner products of the columns of $\mathbf{X}$ with each other, as in (5.3). Hence,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}. \tag{12.7}$$

Furthermore, because $\mathbf{X}$ does not have full column rank, $\mathbf{X}'\mathbf{X}$ is not of full rank.

TABLE 12.2 *X* as a Two-Way Table (Data of Table 12.1)

	Parameters of Model			
Observations	μ	α_1	α_2	α_3
y_{11}	1	1	0	0
y_{12}	1	1	0	0
y_{13}	1	1	0	0
y_{21}	1	0	1	0
y_{22}	1	0	1	0
y_{31}	1	0	0	1

The normal equations (12.6) also involve the vector $\boldsymbol{X}'\boldsymbol{y}$; its elements are the inner products of the columns of $\boldsymbol{X}$ with the vector $\boldsymbol{y}$, and since the only nonzero elements of $\boldsymbol{X}$ are ones, the elements of $\boldsymbol{X}'\boldsymbol{y}$ are certain sums of elements of $\boldsymbol{y}$; for example, from (12.4) and (12.5)

$$\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1\cdot} \\ y_{2\cdot} \\ y_{3\cdot} \end{bmatrix} = \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}. \tag{12.8}$$

This is often the nature of $\boldsymbol{X}'\boldsymbol{y}$ in linear models—a vector of various subtotals of the observations in $\boldsymbol{y}$.

12.2.2 Many Solutions

Whenever $\boldsymbol{X}'\boldsymbol{X}$ is not of full rank, as in (12.4), the normal equations (12.6) cannot be solved with one solitary solution $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ as in Chapter 11. Many solutions are available. To emphasize this we write the normal equations as

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}^{\boldsymbol{0}} = \boldsymbol{X}'\boldsymbol{y}, \tag{12.9}$$

using the symbol $\boldsymbol{\beta}^{\boldsymbol{0}}$ to distinguish those many solutions (12.9) from the solitary solution that exists when $\boldsymbol{X}'\boldsymbol{X}$ has full rank.

Example 12.2 *The normal equations (12.9) for the data of Table 12.1 are from (12.7) and (12.8)*

$$\begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \alpha_3^0 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1\cdot} \\ y_{2\cdot} \\ y_{2\cdot} \end{bmatrix} = \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}. \tag{12.10}$$

Many comments can be made about the detailed form of these equations, for which the reader is referred to Searle (1971, pp. 168 and 232).

12.3 SOLVING THE NORMAL EQUATIONS

We seek a solution for $\boldsymbol{\beta}^{\boldsymbol{0}}$ to the normal equations (12.9): $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}^{\boldsymbol{0}} = \boldsymbol{X}'\boldsymbol{y}$. Since $\boldsymbol{X}'\boldsymbol{X}$ is singular, it has no inverse, and so there are many solutions for $\boldsymbol{\beta}^{\boldsymbol{0}}$, and not just one as there is in regression, in (11.11). This is usually the case with linear models of general nature being discussed here. We therefore obtain a solution for $\boldsymbol{\beta}^{\boldsymbol{0}}$ using a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$.

12.3.1 Generalized Inverses of $X'X$

Let $\boldsymbol{G}$ be a generalized inverse of $\boldsymbol{X'X}$ satisfying

$$\boldsymbol{X'XGX'X} = \boldsymbol{X'X} \tag{12.11}$$

Then, on defining

$$\boldsymbol{X} \text{ as having order } N \times p \text{ with } p < N, \text{ and rank } r_{\boldsymbol{X}} = r \le p, \tag{12.12}$$

Theorem 8.2 gives:

$$\begin{array}{ll} \text{(i)}\quad \boldsymbol{X'XG'X'X} = \boldsymbol{X'X}, & \text{(ii)}\quad \boldsymbol{XGX'X} = \boldsymbol{X}, \\ \text{(iii)}\quad \boldsymbol{XGX'} \text{ is invariant to } \boldsymbol{G}, & \text{(iv)}\quad \boldsymbol{XGX'} \text{ is symmetric.} \end{array} \tag{12.13}$$

Furthermore, on also defining

$$\boldsymbol{H} = \boldsymbol{GX'X}, \tag{12.14}$$

we have,

$$\text{(i) } \boldsymbol{H} \text{ is idempotent, (ii) } r_{\boldsymbol{H}} = r_{\boldsymbol{X}} = r, \text{ and } (iii)\ r(\boldsymbol{H} - \boldsymbol{I}) = p - r, \tag{12.15}$$

(i) follow directly by squaring $\boldsymbol{H}$; (ii) $r_{\boldsymbol{H}} \le r_{\boldsymbol{X}}$, and from the fact that $\boldsymbol{X} = \boldsymbol{XGX'X}$, $r_{\boldsymbol{X}} \le r(\boldsymbol{GX'X}) = r_{\boldsymbol{H}}$, which implies (ii); (iii) is true because $r(\boldsymbol{H} - \boldsymbol{I}) = r(\boldsymbol{I} - \boldsymbol{H})$, and $\boldsymbol{I} - \boldsymbol{H}$ is idempotent; hence, its rank is equal to its trace, that is, $tr(\boldsymbol{I}) - tr(\boldsymbol{H})$, which is equal to $p - rank_{\boldsymbol{H}} = p - r_{\boldsymbol{X}} = p - r$.

12.3.2 Solutions

The above properties indicate that

$$\boldsymbol{\beta^0} = \boldsymbol{GX'y} + (\boldsymbol{H} - \boldsymbol{I})\boldsymbol{z} \tag{12.16}$$

for arbitrary $\boldsymbol{z}$ is a solution for $\boldsymbol{\beta^0}$ of the normal equations (12.9). For our purposes, only one solution is needed, and we take it as $\boldsymbol{GX'y}$ for some generalized inverse $\boldsymbol{G}$. Thus we use

$$\boldsymbol{\beta^0} = \boldsymbol{GX'y} \tag{12.17}$$

as a solution for the normal equations (12.9).

The notation $\boldsymbol{\beta^0}$ in (15.17) for a solution to the normal equations (12.9) emphasizes that, what is derived by solving (12.9) is *only* a solution to the normal equations and *not* an estimator of $\boldsymbol{\beta}$. This fact cannot be over-emphasized. In a general discussion of linear models that are not of full rank, it is essential to realize that what is obtained as a solution of the normal equations is just that, a solution and *nothing more*. It is misleading and in most cases wrong for $\boldsymbol{\beta^0}$ to be termed an estimator of $\boldsymbol{\beta}$. It is true that $\boldsymbol{\beta^0}$ is an estimator of something, but not of $\boldsymbol{\beta}$, and indeed the expression it estimates depends entirely upon which generalized inverse of $\mathbf{X}'\mathbf{X}$ is used in obtaining $\boldsymbol{\beta}^0$. For this reason $\boldsymbol{\beta}^0$ is always referred to as a solution and not an estimator.

Any particular $\boldsymbol{\beta}^0$ depends on which particular generalized inverse of $\mathbf{X}'\mathbf{X}$ is used as $\mathbf{G}$ in calculating $\boldsymbol{\beta}^0 = \mathbf{G}\mathbf{X}'\mathbf{y}$. There is therefore an infinite number of solutions $\boldsymbol{\beta}^0$—except when $\mathbf{X}'\mathbf{X}$ has full rank, whereupon $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1}$ and $\boldsymbol{\beta}^0 = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is then the only solution. Nevertheless, we find that the properties of $\mathbf{G}$ and $\mathbf{H}$ given in (12.13) and (12.15) ensure that this lack of uniqueness in $\boldsymbol{\beta}^0$ is less traumatic than might be anticipated.

Example 12.3 *From (12.10) a generalized inverse of*

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \textit{ is } \mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12.18}$$

for which the solution (12.17) is

$$\boldsymbol{\beta}^0 = \mathbf{G}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix} = \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix}. \tag{12.19}$$

Another generalized inverse and corresponding value of $\boldsymbol{\beta}^0$ *is*

$$\dot{\mathbf{G}} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{5}{6} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & 1\frac{1}{2} \end{bmatrix} \text{ and } \dot{\boldsymbol{\beta}}^0 = \dot{\mathbf{G}}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 86 \\ 14 \\ 0 \\ -54 \end{bmatrix}. \tag{12.20}$$

12.4 EXPECTED VALUES AND VARIANCES

The expected value of $\boldsymbol{\beta}^0 = \mathbf{G}\mathbf{X}'\mathbf{y}$ is

$$E(\boldsymbol{\beta}^0) = \mathbf{G}\mathbf{X}'E(\mathbf{y}) = \mathbf{G}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{H}\boldsymbol{\beta}, \tag{12.21}$$

for $\mathbf{H} = \mathbf{G}\mathbf{X}'\mathbf{X}$. Thus $\boldsymbol{\beta}^0$ has expected value $\mathbf{H}\boldsymbol{\beta}$, and so is an unbiased estimator of $\mathbf{H}\boldsymbol{\beta}$ but *not* of $\boldsymbol{\beta}$.

With $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, the variance-covariance matrix of $\mathbf{y}$ is,

$$\text{var}(\mathbf{y}) = E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'] = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}').$$

and, similarly,

$$\text{var}(\boldsymbol{\epsilon}) = E[(\boldsymbol{\epsilon} - \mathbf{0})(\boldsymbol{\epsilon} - \mathbf{0})'] = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}').$$

Furthermore, we assume that the error terms to be uncorrelated with the variance of each being σ^2, so $\text{var}(\boldsymbol{y}) = \text{var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$. Therefore,

$$\text{var}(\boldsymbol{\beta^0}) = \text{var}(\boldsymbol{GX'y}) = \boldsymbol{GX'}\sigma^2\boldsymbol{IXG'} = \boldsymbol{GX'XG'}\sigma^2. \tag{12.22}$$

Although this is not the analogue of its counterpart $(\boldsymbol{X'X})^{-1}\sigma^2$ in regression as would be $\boldsymbol{G}\sigma^2$, we find that (12.22) does, in fact, create no difficulties in applications.

12.5 PREDICTED y-VALUES

We denote by $\hat{\boldsymbol{y}}$ the vector $\boldsymbol{X\beta^0}$ of predicted y-va1ues corresponding to the vector of observations $\boldsymbol{y}$:

$$\hat{\boldsymbol{y}} = \boldsymbol{X\beta^0} = \boldsymbol{XGX'y}. \tag{12.23}$$

Because $\boldsymbol{G}$ occurs in $\hat{\boldsymbol{y}}$ only in the form $\boldsymbol{XGX'}$, we use (iii) of (12.13) to observe that $\hat{\boldsymbol{y}}$ is invariant to $\boldsymbol{G}$. This means that it does not matter what generalized inverse of $\boldsymbol{X'X}$ is used for $\boldsymbol{G}$: for each and every $\boldsymbol{G}$, the value $\hat{\boldsymbol{y}} = \boldsymbol{XGX'y}$ of (12.23) is the same. This is further emphasized by noting that for any $\boldsymbol{\beta^0}$ of the general form (12.16), that is, for any z in (12.16),

$$\hat{\boldsymbol{y}} = \boldsymbol{X\beta^0} = \boldsymbol{X}[\boldsymbol{GX'y} + (\boldsymbol{H} - \boldsymbol{I})z] = \boldsymbol{XGX'y} + (\boldsymbol{XH} - \boldsymbol{X})z = \boldsymbol{XGX'y}$$

because, by (12.14) and (ii) of (12.13), $\boldsymbol{X\,H} = \boldsymbol{X}$.

Example 12.4 *$\boldsymbol{X}$ is the matrix of 0's and 1's in (12.4). Using $\boldsymbol{\beta^0}$ of (12.19) in $\boldsymbol{X\beta^0}$ of (12.23) therefore gives*

$$\hat{\boldsymbol{y}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 100 \\ 86 \\ 86 \\ 32 \end{bmatrix} \tag{12.24}$$

and using $\dot{\boldsymbol{\beta}}^{\boldsymbol{0}}$ of (12.20) for $\boldsymbol{\beta^0}$ in (12.23) gives the same value for $\hat{\boldsymbol{y}}$:

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 86 \\ 14 \\ 0 \\ -54 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 100 \\ 86 \\ 86 \\ 32 \end{bmatrix}.$$

12.6 ESTIMATING THE ERROR VARIANCE

12.6.1 Error Sum of Squares

Having obtained $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}^0$ as the vector of predicted values corresponding to the vector of observations $\mathbf{y}$, we can now calculate the sum of squares of the deviations of the observed y's from their predicted values. This is the residual (or error) sum of squares, SS_E, and in regression analysis (Section 11.2.2) we found it to be expressible as $SS_E = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$. Essentially the same result holds for the general linear model of this chapter. As before,

$$SS_E = \sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}),$$

so that now

$$\begin{aligned} SS_E &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0) \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')(\mathbf{I} - \mathbf{XGX}')\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}, \end{aligned} \tag{12.25}$$

because from (12.13), $\mathbf{XGX}'$ is both idempotent and symmetric. The occurrence of $\mathbf{G}$ in (12.25) only in the form $\mathbf{XGX}'$ has the same consequence as it does in $\hat{\mathbf{y}}$ of (12.23): the residual sum of squares SS_E is invariant to whatever generalized inverse of $\mathbf{X}'\mathbf{X}$ is used for $\mathbf{G}$, and, equivalently, is invariant to whatever solution of the normal equations $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}'\mathbf{y}$ is used for $\boldsymbol{\beta}^0$. This is especially useful because from (12.25) a computing formula is obtainable that is based on $\boldsymbol{\beta}^0$:

$$\begin{aligned} SS_E &= \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{XGX}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y}. \end{aligned} \tag{12.26}$$

This is exactly the same as in regression: $\mathbf{y}'\mathbf{y}$ is the total sum of squares of the observed y's; and $\boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y}$ is the sum of products of the solutions in $\boldsymbol{\beta}^{0\prime}$ multiplied by the corresponding elements of the right-hand sides of the equations $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}'\mathbf{y}$ from which $\boldsymbol{\beta}^0$ is derived.

Example 12.5 *From (12.4) the total sum of squares, SS_T, is*

$$SS_T = \mathbf{y}'\mathbf{y} = 101^2 + 105^2 + 94^2 + 84^2 + 88^2 + 32^2 = 45,886; \tag{12.27}$$

and using (12.19) in $\boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y}$, with $\mathbf{X}'\mathbf{y}$ being the right-hand side of (12.10), gives

$$\boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y} = 0(504) + 100(300) + 86(172) + 32(32) = 45,816. \tag{12.28}$$

Similarly, using $\boldsymbol{\beta}^{0\prime}$ of (12.20) gives the same value:

$$\boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y} = 86(504) + 14(300) + 0(172) - 54(32) = 45,816. \tag{12.29}$$

Hence, on using either solution (12.19) or (12.20), calculation of SS_E from (12.26) is

$$SS_E = \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}^{\mathbf{0}\prime}\mathbf{X}'\mathbf{y} = 45,886 - 45,816 = 70. \tag{12.30}$$

12.6.2 Expected Value

The mean and variance of $\mathbf{y}$ are, respectively, $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$. And in (12.25), SS_E is a quadratic form in $\mathbf{y}$. Therefore, formula (10.34) can be used to derive the expected value of SS_E as

$$E(SS_E) = E[\mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}] = tr[(\mathbf{I} - \mathbf{XGX}')\mathbf{I}\sigma^2] + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{XGX}')\mathbf{X}\boldsymbol{\beta}.$$

On using the results given in (12.13–12.15) this simplifies to

$$\begin{aligned} E(SS_E) &= \sigma^2[tr(\mathbf{I}) - tr(\mathbf{XGX}')] + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{X} - \mathbf{XGX}'\mathbf{X})\boldsymbol{\beta} \\ &= \sigma^2[N - tr(\mathbf{GX}'\mathbf{X})] + 0 \\ &= \sigma^2(N - r_{\mathbf{H}}) \\ &= \sigma^2(N - r_{\mathbf{X}}). \end{aligned} \tag{12.31}$$

12.6.3 Estimation

Equation (12.31) immediately gives an unbiased estimator of σ^2 as

$$\hat{\sigma}^2 = \frac{SS_E}{N - r_{\mathbf{X}}}. \tag{12.32}$$

Again we see a similarity with regression. Only now, the importance of using $r_{\mathbf{X}}$ in the expectation is clear, because $\mathbf{X}$ is not of full column rank and its rank is therefore not equal to its number of columns. In fact, the rank of $\mathbf{X}$ depends on the nature of the data available.

Example 12.6

$$\hat{\sigma}^2 = \frac{70}{6-3} = 23\tfrac{1}{3}. \tag{12.33}$$

12.7 PARTITIONING THE TOTAL SUM OF SQUARES

Partitioning the total sum of squares as done for regression in Chapter 11 takes place in the same way for the general linear model being dealt with in this chapter. One difference is that corrected sums of squares and products of x-variables are of neither use nor interest (because every x-variable is just 0 or 1, representing the incidence of a term of the model in the data). Nevertheless, use is still made of $SS_{Tot} = \mathbf{y}'\mathbf{y} - N\bar{y}^2$, the corrected sum of squares of the y-observations. The three forms of partitioning the sums of squares are shown in Table 12.3.

TABLE 12.3 Partitioning Sums of Squares

	$N\bar{y}^2$	
$SS_{reg} = \mathbf{y}'\mathbf{XGX}'\mathbf{y}$	$SS_{Reg} = \mathbf{y}'\mathbf{XGX}'\mathbf{y} - N\bar{y}^2$	$SS_{Reg} = \mathbf{y}'\mathbf{XGX}'\mathbf{y} - N\bar{y}^2$
$SS_E = \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}$	$SS_E = \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}$	$SS_E = \mathbf{y}'(\mathbf{I} - \mathbf{XGX}')\mathbf{y}$
$SS_T = \mathbf{y}'\mathbf{y}$	$SS_T = \mathbf{y}'\mathbf{y}$	$SS_{Tot} = \mathbf{y}'\mathbf{y} - N\bar{y}^2$

The three columns in Table 12.3 correspond to the three partitionings shown in (11.23) and (11.27). The first column shows

$$SS_{reg} = SS_T - SS_E = \mathbf{y}'\mathbf{XGX}'\mathbf{y} = \boldsymbol{\beta}^{0\prime}\mathbf{X}'\mathbf{y}, \tag{12.34}$$

the sum of squares attributable to fitting the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, similar to the sum of squares for regression in Section 11.1.1. In the second column, $N\bar{y}^2$ is the sum of squares due to fitting a general mean, and

$$SS_{Reg} = SS_{reg} - N\bar{y}^2 \tag{12.35}$$

is the sum of squares for fitting the model, corrected for the mean. The third column is identical to the second except that $N\bar{y}^2$ has been deleted from the body of the table and subtracted from SS_T to give

$$SS_{Tot} = \sum_{i,j} y_{ij}^2 - N\bar{y}^2 \tag{12.36}$$

as the total sum of squares corrected for the mean. In all three columns the error sum of squares, SS_E, is the same.

Table 12.3 forms the basis of traditional ANOVA tables, as is shown in Section 12.8.

Example 12.7 *With $\bar{y} = 504/6$ from Table 12.1, we have $N\bar{y}^2 = 42,336$; and with $SS_T = 45,886$ and $SS_E = 70$ from (12.27) and (12.30), respectively, Table 12.3 for the example is as shown in Table 12.4.*

12.8 ANALYSIS OF VARIANCE

Analysis of variance is based on partitioning sums of squares in the manner of Tables 12.3 and 12.4. The right-most column of those tables is the partitioning seen in most analysis of variance tables, with the total sum of squares being that corrected for the mean,

TABLE 12.4 Partitioning Sums of Squares as in Table 12.3 (Data of Table 12.1)

	$N\bar{y}^2 = 42,336$	
$SS_{reg} = 45,816$	$SS_{Reg} = 3480$	$SS_{Reg} = 3480$
$SS_E = 70$	$SS_E = 70$	$SS_E = 70$
$SS_T = 45,886$	$SS_T = 45,886$	$SS_{Tot} = 3550$

TABLE 12.5 Analysis of Variance

Source of Variation	DF[a]	SS	MS	F-Statistic
(a) Based on Table 12.3				
Model, a.f.m.[b]	$r_{\mathbf{X}} - 1$	SS_{Reg}	$MS_{Reg} = \frac{SS_{Reg}}{r_{\mathbf{X}}-1}$	$\frac{MS_{Reg}}{MS_E}$
Error	$N - r_{\mathbf{X}}$	SS_E	$MS_E = \frac{SS_E}{N-r_{\mathbf{X}}}$	
Total	$N - 1$	SS_{Tot}		
(b) For Data of Table 12.1, Based on Table 12.4				
Model a.f.m.[b]	2	3480	$\frac{3480}{2} = 1740$	$\frac{1740}{23\frac{1}{3}} = 74\frac{3}{7}$
Error	3	70	$\frac{70}{3} = 23\frac{1}{3}$	
Total	5	3550		

[a] DF = degrees of freedom.
[b] a.f.m. = adjusted for mean.

$SS_{Tot} = \boldsymbol{y}'\boldsymbol{y} - N\bar{y}^2$. In contrast, we will concentrate attention on the middle column of Tables 12.3 and 12.4, which has the term $N\bar{y}^2$ in it, as well as SS_{Reg} and SS_E. The analysis of variance based on this column is shown in Table 12.5, first in terms of the sums of squares in Table 12.3, and then for the data, using Table 12.4.

Table 12.5 is just a summary of calculations. To utilize the F-statistics therein, in the usual manner, we assume that the observations in $\boldsymbol{y}$ are normally distributed with mean $\boldsymbol{X\beta}$ and variance-covariance matrix var($\boldsymbol{y}$) = $\sigma^2\boldsymbol{I}$ and write

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}). \tag{12.37}$$

Equivalently, the assumptions can be stated with respect to the error vector, $\boldsymbol{\epsilon}$, as

$$\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}). \tag{12.38}$$

Under these assumptions we have the following distributional properties:

(i) $\frac{1}{\sigma^2}SS_E \sim \chi^2_{N-r_X}$.
(ii) $\frac{1}{\sigma^2}SS_{reg} \sim \chi^2_{r_X}(\theta)$, where $\theta = \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}$.
(iii) SS_E and SS_{reg} are independent.

Properties (i) and (ii) result from applying Theorem 10.2 since SS_{reg} and SS_E are quadratic forms in $\boldsymbol{y}$ with corresponding matrices, $\boldsymbol{I} - \boldsymbol{XGX}'$ and $\boldsymbol{XGX}'$ that are idempotent of ranks $N - r_X$ and r_X, respectively. The noncentrality parameter for $\frac{1}{\sigma^2}SS_E$ is zero and the one for $\frac{1}{\sigma^2}SS_{reg}$ is

$$\begin{aligned}\theta &= \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'[\boldsymbol{XGX}']\boldsymbol{X\beta} \\ &= \frac{1}{\sigma^2}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}.\end{aligned}$$

Property (iii) is true by Theorem 10.3 and the fact that

$$[\boldsymbol{XGX}'](\sigma^2\boldsymbol{I})[\boldsymbol{I} - \boldsymbol{XGX}'] = \boldsymbol{0}.$$

These results establish that the F-statistics in Table 12.5 have F-distributions with degrees of freedom corresponding to those of the mean squares in their numerator and denominator.

We look upon the F-ratio, $\frac{MS_{Reg}}{MS_E}$, as providing a test of the model $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ over and above the mean. When this F-ratio is significant we conclude that the model satisfactorily accounts for variation in the y-variable. This is not to be taken as evidence that all elements of $\boldsymbol{\beta}$ other than μ are nonzero but only that at least one of them, or one linear combination of them, may be.

Example 12.8 *The particular hypothesis tested by the F-ratio,* $\frac{MS_{Reg}}{MS_E}$, *in Table 12.5b is as follows:*

$$\frac{MS_{Reg}}{MS_E} \text{ tests } H : \alpha_1 = \alpha_2 = \alpha_3. \tag{12.39}$$

The latter is the familiar hypothesis of equality of the effects of type of plant on yield.

12.9 THE $R(\cdot|\cdot)$ NOTATION

In Section 11.9, the notation $R(\boldsymbol{\beta_1} | \boldsymbol{\beta_2})$ was introduced in the context of regression as the sum of squares due to fitting $E(\boldsymbol{y}) = \boldsymbol{X}_1\boldsymbol{\beta_1} + \boldsymbol{X}_2\boldsymbol{\beta_2}$ after fitting $E(\boldsymbol{y}) = \boldsymbol{X}_2\boldsymbol{\beta_2}$. This is now applied to the model $E(y_{ij}) = \mu + \alpha_i$ of (12.2), denoting the sum of squares for this as $R(\mu, \alpha)$. It is, of course, SS_{reg} in the analysis of variance:

$$R(\mu, \alpha) = SS_{reg}. \tag{12.40}$$

Similarly, $R(\mu)$ denotes the sum of squares due to fitting $E(y_{ij}) = \mu$ and so

$$R(\mu) = N\bar{y}^2 \tag{12.41}$$

Hence, in combination with (12.35) and analogous to (11.84),

$$SS_{Reg} = R(\mu, \alpha) - R(\mu) = R(\alpha | \mu). \tag{12.42}$$

In this way $R(\alpha | \mu)$ indicates by its very symbolism that it is the sum of squares due to having α's in the model, over and above having just μ; i.e., that it is the sum of squares due to fitting $E(y_{ij}) = \mu + \alpha_i$ over and above that due to fitting $E(y_{ij}) = \mu$. This proves to be a useful notation in models of this nature that have more than one factor, as indicated in Section 12.12.

12.10 ESTIMABLE LINEAR FUNCTIONS

We have seen earlier in this chapter that $\boldsymbol{\beta}^0$ is just a solution to the normal equations (12.9), but cannot be considered as an estimator of $\boldsymbol{\beta}$ due to its non-uniqueness. Hence, not every linear function of $\boldsymbol{\beta}$ of the form $\boldsymbol{a}'\boldsymbol{\beta}$ can be estimated. There are, however, certain conditions under which $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable.

Definition 12.1 *Let $\boldsymbol{a}$ is a constant vector. The linear function, $\boldsymbol{a}'\boldsymbol{\beta}$, is said to be* estimable *if there exists a linear function of $\boldsymbol{y}$, the vector of observations in (12.1), of the form $\boldsymbol{t}'\boldsymbol{y}$ such that $E(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{\beta}$.*

The necessary and sufficient condition for the estimability of $\boldsymbol{a}'\boldsymbol{\beta}$ is given by the following theorem.

Theorem 12.1 *The linear function $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{a}'$ belongs to the row space of $\boldsymbol{X}$ in (12.1), that is, $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}$ for some vector $\boldsymbol{t}$.*

Proof. If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then by Definition 12.1 there exists a vector $\boldsymbol{t}$ such that $E(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{\beta}$. Hence, for all $\boldsymbol{\beta} \in R^p$, where p is the number of columns of $\boldsymbol{X}$, we have

$$\boldsymbol{t}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{\beta}, \tag{12.43}$$

It follows that

$$\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}. \tag{12.44}$$

This indicates that $\boldsymbol{a}'$ is a linear combination of the rows of $\boldsymbol{X}$ and must therefore belong to the row space of $\boldsymbol{X}$. Vice versa, if $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}$, for some $\boldsymbol{t}$, then $E(\boldsymbol{t}'\boldsymbol{y}) = \boldsymbol{t}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{\beta}$, which makes $\boldsymbol{a}'\boldsymbol{\beta}$ estimable. ■

Another necessary and sufficient condition for the estimability of $\boldsymbol{a}'\boldsymbol{\beta}$ is given by the following corollary:

Corollary 12.1 *The linear function $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable if and only if the matrix*

$$\boldsymbol{X}_a = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{a}' \end{bmatrix}$$

has the same rank as that of $\boldsymbol{X}$, that is, if and only if the numbers of nonzero eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}_a{}'\boldsymbol{X}_a$ are the same.

Since $\boldsymbol{X}$ is of rank r, then from Theorem 12.1 we can conclude that the number of linearly independent estimable functions of $\boldsymbol{\beta}$ is equal to r (see also Exercise 12.1).

12.10.1 Properties of Estimable Functions

Estimable linear functions have several interesting features in the sense that their estimation and tests of significance are carried out in much the same way as in the case of full-rank linear models in Chapter 11. Their properties are described by the next three theorems.

Theorem 12.2 *If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$, where $\boldsymbol{\beta}^{\mathbf{0}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$, is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^{-}$.*

Proof. If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then by Theorem 12.1, $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}$ for some vector $\boldsymbol{t}$. Hence,

$$\begin{aligned}\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}} &= \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}.\end{aligned}$$

Invariance of $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ follows from the fact that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}$ is invariant to the choice of the generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$ (see Property (iii) in Section 12.3.1). ■

The uniqueness of $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ for a given $\boldsymbol{y}$, when $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, makes $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ a full-fledged estimator of $\boldsymbol{a}'\boldsymbol{\beta}$, which was not the case with $\boldsymbol{\beta}^{\mathbf{0}}$.

Theorem 12.3 **(The Gauss–Markov Theorem)** *Suppose that ϵ in model (12.1) has a zero mean vector and a variance-covariance matrix given by $\sigma^2\boldsymbol{I}$. If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}} = \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{a}'\boldsymbol{\beta}$.*

Proof. The proof is very similar to the proof of the Gauss–Markov Theorem in Chapter 11 (Theorem 11.1). Details of the proof are left to the reader. ■

Theorem 12.4 *The linear function $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{a}' = \boldsymbol{a}'\boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}$.*

Proof. If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}$ for some vector $\boldsymbol{t}$. Hence, by Theorem 8.2(ii), $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{a}'\boldsymbol{H}$. Conversely, if $\boldsymbol{a}' = \boldsymbol{a}'\boldsymbol{H}$, then $\boldsymbol{a}' = \boldsymbol{a}'\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{t}'\boldsymbol{X}$, where $\boldsymbol{t}' = \boldsymbol{a}'\boldsymbol{G}\boldsymbol{X}'$, which implies that $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable. ■

Theorem 12.5 *The linear function $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{a}' = \boldsymbol{w}'\boldsymbol{H}$ for an arbitrary vector $\boldsymbol{w}$.*

Proof. If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then by Theorem 8.2(ii), $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X} = \boldsymbol{t}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{w}'\boldsymbol{H}$, where $\boldsymbol{w}' = \boldsymbol{t}'\boldsymbol{X}$. Conversely, if $\boldsymbol{a}' = \boldsymbol{w}'\boldsymbol{H}$, then $\boldsymbol{a}' = \boldsymbol{w}'\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}$, which indicates that $\boldsymbol{a}'$ belongs to the row space of $\boldsymbol{X}$ and therefore $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable. ■

If $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then its BLUE is $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}} = \boldsymbol{w}'\boldsymbol{H}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{w}'\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{w}'\boldsymbol{G}\boldsymbol{X}'\boldsymbol{y}$, by applying Theorem 8.2(ii).

Theorem 12.6 *Suppose that $\boldsymbol{\epsilon}$ in model (12.1) is distributed as $N(\mathbf{0}, \sigma^2\boldsymbol{I})$, and that $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable. Then,*

(a) *$\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ is normally distributed with mean $\boldsymbol{a}'\boldsymbol{\beta}$ and variance $\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{a}\,\sigma^2$, where $\boldsymbol{\beta}^{\mathbf{0}}$ is given by (12.17).*

(b) *$\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ and MS_E are independently distributed, where MS_E is the error mean square,*

$$MS_E = \frac{1}{n - r_{\boldsymbol{X}}}\boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']\boldsymbol{y} \tag{12.45}$$

Proof.

(a) $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}} = \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$, for some vector $\boldsymbol{t}$, is normally distributed since $\boldsymbol{y}$ is normally distributed. Its mean is

$$\begin{aligned} E(\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}) &= \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{t}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{a}'\boldsymbol{\beta}, \end{aligned}$$

and its variance is

$$\begin{aligned} \text{var}(\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}) &= \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'(\sigma^2\boldsymbol{I})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{t} \\ &= \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{t}\,\sigma^2 \\ &= \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{a}\,\sigma^2. \end{aligned}$$

(b) $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$ and MS_E are independent by applying Theorem 10.4 and knowing that

$$\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'(\sigma^2\boldsymbol{I})[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = \mathbf{0}'.$$

■

12.10.2 Testable Hypotheses

Using the properties described in Section 12.10.1 we can now derive tests and confidence intervals concerning estimable linear functions of $\boldsymbol{\beta}$.

Definition 12.2 *The hypothesis $H_0 : \boldsymbol{K\beta} = \boldsymbol{m}$ is said to be* testable *if the elements of $\boldsymbol{K\beta}$ are estimable, where $\boldsymbol{K}$ is a matrix of order $s \times p$ and rank s ($\leq r$, the rank of $\boldsymbol{X}$ in model (12.1)), and $\boldsymbol{m}$ is a constant vector.*

Lemma 12.1 *The hypothesis $H_0 : \boldsymbol{K\beta} = \boldsymbol{m}$ is testable if and only if there exists a matrix $\boldsymbol{S}$ of order $s \times p$ such that $\boldsymbol{K} = \boldsymbol{SX}'\boldsymbol{X}$.*

Proof. If $\boldsymbol{K\beta}$ is testable, then the rows of $\boldsymbol{K}$ must belong to the row space of $\boldsymbol{X}$. Thus, there exists a matrix $\boldsymbol{T}$ such that $\boldsymbol{K} = \boldsymbol{TX}$. Hence, $\boldsymbol{K} = \boldsymbol{TX}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{SX}'\boldsymbol{X}$, where

$\boldsymbol{S} = \boldsymbol{TX}(\boldsymbol{X}'\boldsymbol{X})^-$. Vice versa, if $\boldsymbol{K} = \boldsymbol{SX}'\boldsymbol{X}$ for some matrix $\boldsymbol{S}$, then any row of $\boldsymbol{K}$ is a linear combination of the rows of $\boldsymbol{X}$ implying estimability of $\boldsymbol{K\beta}$ and hence testability of H_0. ■

12.10.3 Development of a Test Statistic for H_0

Suppose that $\boldsymbol{K\beta}$ is estimable, where $\boldsymbol{K}$ is $s \times p$ of rank s $(\leq r)$. The best linear unbiased estimator of $\boldsymbol{K\beta}$ is $\boldsymbol{K}\hat{\boldsymbol{\beta}} = \boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y}$, assuming that $\boldsymbol{\epsilon}$ in (12.1) has a zero mean and a variance-covariance matrix $\sigma^2\boldsymbol{I}$. The variance-covariance matrix of $\boldsymbol{K}\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} \text{var}(\boldsymbol{K}\hat{\boldsymbol{\beta}}) &= \boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{K}'\,\sigma^2 \\ &= \boldsymbol{SX}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{XS}'\,\sigma^2 \\ &= \boldsymbol{SX}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{XS}'\,\sigma^2 \\ &= \boldsymbol{SX}'\boldsymbol{XS}'\,\sigma^2, \end{aligned} \tag{12.46}$$

where $\boldsymbol{S}$ is the matrix described in Lemma 12.1. Furthermore, $\text{var}(\boldsymbol{K}\hat{\boldsymbol{\beta}})$ can also be expressed as

$$\begin{aligned} \text{var}(\boldsymbol{K}\hat{\boldsymbol{\beta}}) &= \boldsymbol{TX}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{T}'\,\sigma^2 \\ &= \boldsymbol{TX}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{T}'\,\sigma^2 \\ &= \boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{K}'\,\sigma^2, \end{aligned} \tag{12.47}$$

where $\boldsymbol{T}$ is a matrix such that $\boldsymbol{K} = \boldsymbol{TX}$. It can be seen that $\text{var}(\boldsymbol{K}\hat{\boldsymbol{\beta}})$ is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^-$ and is a nonsingular matrix. The second assertion is true because $\boldsymbol{K}$ is of full row rank s and

$$s = rank(\boldsymbol{K}) \leq rank(\boldsymbol{SX}') \leq rank(\boldsymbol{S}) \leq s,$$

since $\boldsymbol{S}$ has s rows. Hence, $rank(\boldsymbol{SX}') = s$ and the matrix $\boldsymbol{SX}'\boldsymbol{XS}'$ is therefore of full rank, which implies that $\text{var}(\boldsymbol{K}\hat{\boldsymbol{\beta}})$ is nonsingular by (12.46).

A test statistic concerning $H_0 : \boldsymbol{K\beta} = \boldsymbol{m}$ versus $H_a : \boldsymbol{K\beta} \neq \boldsymbol{m}$ can now be obtained, assuming that $\boldsymbol{K\beta}$ is estimable and $\boldsymbol{\epsilon}$ in model (12.1) is distributed as $N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. Since

$$\boldsymbol{K}\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{K\beta},\, \boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{K}'\,\sigma^2],$$

then, under H_0,

$$F = \frac{(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{m})'[\boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{K}']^{-1}(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{m})}{s\,MS_E} \tag{12.48}$$

has the F-distribution with s and $n - r$ degrees of freedom, where MS_E is the error mean square in (12.45). This is true because $\boldsymbol{K}\hat{\boldsymbol{\beta}}$ is independent of MS_E by Theorem 12.6 and $\frac{1}{\sigma^2}SS_E \sim \chi^2_{n-r}$. The null hypothesis can therefore be rejected at the α - level if $F \geq F_{\alpha,s,n-r}$.

The power of this test under the alternative hypothesis $H_a : \boldsymbol{K\beta} = \boldsymbol{m}_a$, where $\boldsymbol{m}_a$ is a given constant vector different from $\boldsymbol{m}$, is

$$\text{Power} = P[F \geq F_{\alpha,s,n-r} | H_a : \boldsymbol{K\beta} = \boldsymbol{m}_a].$$

Under H_a, F has the noncentral F-distribution $F_{s,n-r}(\theta)$ with the noncentrality parameter

$$\theta = \frac{1}{\sigma^2}(\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{K}']^{-1}(\boldsymbol{m}_a - \boldsymbol{m}). \tag{12.49}$$

Hence, the power is given by

$$\text{Power} = P[F_{s,n-r}(\theta) \geq F_{\alpha,s,n-r}]$$

A confidence region on $\boldsymbol{K\beta}$ can also be obtained on the basis of the aforementioned F-distribution. We have that,

$$\frac{(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{K\beta})'[\boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{K}']^{-1}(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{K\beta})}{s\, MS_E}$$

has the F-distribution with s and $n - r$ degrees of freedom. It follows that the $(1 - \alpha)100\%$ confidence region on $\boldsymbol{K\beta}$ is of the form

$$\frac{(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{K\beta})'[\boldsymbol{K}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{K}']^{-1}(\boldsymbol{K}\hat{\boldsymbol{\beta}} - \boldsymbol{K\beta})}{s\, MS_E} \leq F_{\alpha,s,n-r}. \tag{12.50}$$

In particular, when $\boldsymbol{K} = \boldsymbol{k}'$, a test statistic concerning $H_0 : \boldsymbol{k}'\boldsymbol{\beta} = m$ versus $H_a : \boldsymbol{k}'\boldsymbol{\beta} \neq m$ is

$$t = \frac{\boldsymbol{k}'\hat{\boldsymbol{\beta}} - m}{[\boldsymbol{k}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{k}\, MS_E]^{1/2}}, \tag{12.51}$$

which, under H_0, has the t-distribution with $n - r$ degrees of freedom. In this case, the $(1 - \alpha)100\%$ confidence interval on $\boldsymbol{k}'\boldsymbol{\beta}$ is given by

$$\boldsymbol{k}'\hat{\boldsymbol{\beta}} \pm [\boldsymbol{k}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{k}\, MS_E]^{1/2}\, t_{\frac{\alpha}{2},n-r}. \tag{12.52}$$

Example 12.9 *Consider the one-way model,*

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i,$$

where α_i is a fixed unknown parameter, and the ϵ_{ij}'s are independently distributed as $N(0, \sigma^2)$. This model can be written as in (12.1) with $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_k)'$ and

$$\boldsymbol{X} = [\boldsymbol{1}_{n_.} : \oplus_{i=1}^{k} \boldsymbol{1}_{n_i}],$$

where $n_. = \sum_{i=1}^{k} n_i$ *and* $\mathbf{1}_{n_i}$ *is a vector of ones of order* $n_i \times 1$ $(i = 1, 2, \ldots, k)$*. The matrix* $\boldsymbol{X}$ *is of rank k and its row space is spanned by the k vectors,*

$$(1, 1, 0, \ldots, 0),\ (1, 0, 1, \ldots, 0), \ldots, (1, 0, 0, \ldots, 0, 1),$$

which are linearly independent. On this basis we have the following results:

Result 1 $\mu + \alpha_i$ $(i = 1, 2, \ldots, k)$ *form a basis for all estimable linear functions of* $\boldsymbol{\beta}$*. Hence,* $\alpha_{i_1} - \alpha_{i_2}$ *is estimable for* $i_1 \neq i_2$*.*

Result 2 μ *is nonestimable.*

To show this result, let us write μ *as* $\boldsymbol{a}'\boldsymbol{\beta}$*, where* $\boldsymbol{a}' = (1, 0, 0, \ldots, 0)$*. If* μ *is estimable, then* $\boldsymbol{a}'$ *must belong to the row space of* $\boldsymbol{X}$*, that is,* $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X}$*, for some vector* $\boldsymbol{t}$*. In this case, we have*

$$\boldsymbol{t}'\mathbf{1}_{n_.} = 1 \tag{12.53}$$

$$\boldsymbol{t}' \oplus_{i=1}^{k} \mathbf{1}_{n_i} = \mathbf{0}'. \tag{12.54}$$

Equality (12.54) shows that $\boldsymbol{t}$ *is orthogonal to the columns of* $\oplus_{i=1}^{k}\mathbf{1}_{n_i}$*. It must therefore be orthogonal to* $\mathbf{1}_{n_.}$*, which is the sum of the columns of* $\oplus_{i=1}^{k}\mathbf{1}_{n_i}$*. This contradicts equality (12.53). We conclude that* μ *is nonestimable.*

Result 3 *The best linear unbiased estimator (BLUE) of* $\alpha_{i_1} - \alpha_{i_2}$*,* $i_1 \neq i_2$*, is* $\bar{y}_{i_1.} - \bar{y}_{i_2.}$*, where* $\bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$*,* $i = 1, 2, \ldots, k$*.*

To show this, let us first write $\mu + \alpha_i$ *as* $\boldsymbol{a}_i'\boldsymbol{\beta}$*, where* $\boldsymbol{a}_i'$ *is a vector with* $k+1$ *elements; its first element is 1 and the element corresponding to* α_i *is also 1. Since* $\mu + \alpha_i$ *is estimable, its BLUE is* $\boldsymbol{a}_i'\boldsymbol{\beta}^{\mathbf{0}}$*, where*

$$\begin{aligned}\boldsymbol{\beta}^{\mathbf{0}} &= (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \\ &= \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} y_{..} \\ y_{1.} \\ \cdot \\ \cdot \\ \cdot \\ y_{k.} \end{bmatrix},\end{aligned} \tag{12.55}$$

where $y_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$*,* $y_{i.} = n_i\bar{y}_{i.}$*, and* $\boldsymbol{D} = diag(n_1^{-1}, n_2^{-1}, \ldots, n_k^{-1})$*. From (12.55) it follows that* $\mu^{\mathbf{0}} = 0$*,* $\alpha_i^{\mathbf{0}} = \bar{y}_{i.}$*,* $i = 1, 2, \ldots, k$*. Hence,* $\boldsymbol{a}_i'\boldsymbol{\beta}^{\mathbf{0}} = \bar{y}_{i.}$*. Consequently, the BLUE of* $\alpha_{i_1} - \alpha_{i_2}$ *is* $\bar{y}_{i_1.} - \bar{y}_{i_2.}$*,* $i_1 \neq i_2$*. Its variance is*

$$var(\bar{y}_{i_1.} - \bar{y}_{i_2.}) = \left(\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}\right)\sigma^2.$$

The $(1-\alpha)100\%$ *confidence interval on* $\alpha_{i_1} - \alpha_{i_2}$ *is then given by*

$$\bar{y}_{i_1.} - \bar{y}_{i_2.} \pm \left[\left(\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}\right)MS_E\right]^{1/2} t_{\alpha/2, n_.-k}$$

12.11 CONFIDENCE INTERVALS

Since it is only estimable functions that have estimators (BLUE's) that are invariant to the solution of the normal equations, they are the only functions for which establishing confidence intervals is valid. On assuming $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ we then have, that for an estimable function $\boldsymbol{a}'\boldsymbol{\beta}$ its BLUE is $\boldsymbol{a}'\boldsymbol{\beta}^{\boldsymbol{0}} \sim N(\boldsymbol{a}'\boldsymbol{\beta}, \boldsymbol{a}'\boldsymbol{Ga}\,\sigma^2)$. Therefore, akin to Section 11.7.2, a $(1-\alpha)100\%$ confidence interval on $\boldsymbol{a}'\boldsymbol{\beta}$ is

$$\boldsymbol{a}'\boldsymbol{\beta}^{\boldsymbol{0}} \pm t_{\frac{\alpha}{2},n-r}\sqrt{\boldsymbol{a}'\boldsymbol{Ga}MS_E},$$

where n is the number observations and r is the rank of $\boldsymbol{X}$.

12.12 SOME PARTICULAR MODELS

The general results of the preceding sections are now used in three particular models of widespread application.

12.12.1 The One-Way Classification

This is the model (12.2), $E(y_{ij}) = \mu + \alpha_i$, of the example used to this point. The general form of the normal equations is

$$\begin{bmatrix} N & n_1 & n_2 & \cdots & n_k \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & & & & \\ n_k & 0 & \cdots & 0 & n_k \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \vdots \\ \alpha_k^0 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{k.} \end{bmatrix} \tag{12.56}$$

where k is the number of groups or classes, corresponding to the three types of plant in Table 12.1, where

$$N = \sum_{i=1}^{k} n_i, \quad y_{i.} = \sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad \bar{y}_{i.} = y_{i.}/n_i \quad \text{and} \quad y_{..} = \sum_{i=1}^{k} y_{i.}.$$

The solution to the normal equations corresponding to (12.19) is

$$\mu^0 = 0 \quad \text{and} \quad \alpha_i^0 = \bar{y}_{i.} \tag{12.57}$$

The basic estimable function is $\mu + \alpha_i$, and its BLUE is

$$\widehat{\mu + \alpha_i} = \mu^0 + \alpha_i^0 = 0 + \bar{y}_{i.} = \bar{y}_{i.}. \tag{12.58}$$

Any linear combination of the $(\mu + \alpha_i)$-expressions is also estimable, with its BLUE being that same linear combination of the $\bar{y}_{i.}$'s.

TABLE 12.6 Analysis of Variance for the One-Way Classification

Source of Variation	DF	SS	F-Statistic
Mean	1	$R(\mu) = N\bar{y}^2$	
Classes, a.f.m.[a]	$k-1$	$R(\alpha\|\mu) = R(\mu, \alpha) - N\bar{y}^2$	$F(\alpha\|\mu) = \dfrac{R(\alpha\|\mu)/(k-1)}{\text{SSE}/(N-k)}$
Error	$N-k$	$SS_E = \mathbf{y}'\mathbf{y} - R(\mu, \alpha)$	
Total	N	$SS_T = \mathbf{y}'\mathbf{y}$	

[a]a.f.m. = adjusted for mean.

A feature of the model is its analysis of variance. Using (12.56) and (12.57) in (12.34) gives (12.40) as

$$R(\mu, \alpha) = SS_{reg} = 0(y_{..}) + \sum_{i=1}^{k} \bar{y}_{i.}y_{i.} = \sum_{i=1}^{k} n_i\bar{y}_{i.}^2 = \sum_{i=1}^{k} y_{i.}{}^2/n_i \tag{12.59}$$

and, as usual, $R(\mu) = N\bar{y}_{..}^2$. Using these and SS_{Reg} of (12.42) in the first part of Table 12.5 gives the analysis of variance shown in Table 12.6.Part (b) of Table 12.5 is an example of Table 12.6. The F-statistic, $F(\alpha|\mu)$, tests the hypothesis $H : \alpha_1 = \alpha_2 = \cdots = \alpha_k$. Full description of this analysis of variance and of other details of this model are to be found in (Searle 1971, Section 7.2).

12.12.2 Two-Way Classification, No Interactions, Balanced Data

Countless experiments are undertaken each year in agriculture and the plant sciences to investigate the effect on growth and yield of various fertilizer treatments applied to different varieties of a species. Suppose we have data from six plants, representing three varieties being tested in combination with two fertilizer treatments. Although the experiment would not necessarily be conducted by growing the plants in two rows of three plants each, it is convenient to visualize the data as in Table 12.7. The entries in the table are such that y_{ij} represents the yield of the plant of variety i that received fertilizer treatment j. If μ is a general mean for plant yield, if α_i represents the effect on yield due to variety i and β_j the effect due to treatment j, the equation of the linear model for y_{ij} is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \tag{12.60}$$

where ϵ_{ij} is a random error term peculiar to y_{ij}. This is the model for the two-way classification without interaction. It is a straightforward extension of the model given in (12.2) for

TABLE 12.7 Yields of Plants

	Treatment		
Variety	1	2	Total
1	y_{11}	y_{12}	$y_{1.}$
2	y_{21}	y_{22}	$y_{2.}$
3	y_{31}	y_{32}	$y_{3.}$
Total	$y_{.1}$	$y_{.2}$	$y_{..}$

TABLE 12.8 Estimators of Estimable Functions (In The Two-Way Classification, No Interaction Model, Balanced Data)

Values of w's				Estimable Function, $\mathbf{a}'\boldsymbol{\beta}$	BLUE, $\mathbf{a}'\boldsymbol{\beta}^0$
w_1	w_2	w_3	w_4	Equation (12.62)	Equation (12.63)
1	−1	0	0	$\alpha_1 - \alpha_2$	$\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$
1	0	−1	0	$\alpha_1 - \alpha_3$	$\bar{y}_{1\cdot} - \bar{y}_{3\cdot}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{2}$	$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3) + \frac{1}{2}(\beta_1 + \beta_2)$	$\bar{y}_{\cdot\cdot}$
0	0	0	1	$\beta_1 - \beta_2$	$\bar{y}_{\cdot 1} - \bar{y}_{\cdot 2}$

the one-way classification. As before, it is assumed that the ϵ's are independent with zero means and variance σ^2.

On writing each observation in terms of (12.60) we get

$$\begin{array}{llllllll}
y_{11} = \mu + \alpha_1 & & & +\beta_1 & & +\epsilon_{11} \\
y_{12} = \mu + \alpha_1 & & & & +\beta_2 & +\epsilon_{12} \\
y_{21} = \mu & +\alpha_2 & & +\beta_1 & & +\epsilon_{21} \\
y_{22} = \mu & +\alpha_2 & & & +\beta_2 & +\epsilon_{22} \\
y_{31} = \mu & & +\alpha_3 & +\beta_1 & & +\epsilon_{31} \\
y_{32} = \mu & & +\alpha_3 & & +\beta_2 & +\epsilon_{32}
\end{array}$$

which can easily be written in matrix form as

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \boldsymbol{\epsilon} \tag{12.61}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are the vectors of observations and error terms, respectively. With $\mathbf{X}$ representing the matrix of 0's and 1's and $\boldsymbol{\beta}$ the vector of parameters, this equation is the familiar form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Accordingly, we find

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \\ y_{3\cdot} \\ y_{\cdot 1} \\ y_{\cdot 2} \end{bmatrix}, \quad \text{and} \quad \mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 \end{bmatrix}$$

for which a generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$\mathbf{G} = \frac{1}{6} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 1 & -2 & 0 \\ 0 & 1 & 4 & 1 & -2 & 0 \\ 0 & 1 & 1 & 4 & -2 & 0 \\ 0 & -2 & -2 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

With $\mathbf{w}' = (w_0, \quad w_1, \quad w_2, \quad w_3, \quad w_4, \quad w_5)$ we find from Theorems 12.4 and 12.5 after calculating $\mathbf{H} = \mathbf{G}\mathbf{X}'\mathbf{X}$ that for any values of the $\mathbf{w}$'s estimable functions are

$$\begin{aligned} \mathbf{a}'\boldsymbol{\beta} = \mathbf{w}'\mathbf{H}\boldsymbol{\beta} = (w_1 + w_2 + w_3)\mu + w_1\alpha_1 + w_2\alpha_2 + w_3\alpha_3 \\ + w_4\beta_1 + (w_1 + w_2 + w_3 - w_4)\beta_2 \end{aligned} \tag{12.62}$$

and their BLUE's are, from Section 12.10.1,

$$\begin{aligned} \mathbf{a}'\boldsymbol{\beta}^0 &= \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{y} \\ &= w_1(4y_{1.} + y_{2.} + y_{3.} - 2y_{.1})/6 + w_2(y_{1.} + 4y_{2.} + y_{3.} - 2y_{.1})/6 \\ &\quad + w_3(y_{1.} + y_{2.} + 4y_{3.} - 2y_{.1})/6 - w_4(2y_{1.} + 2y_{2.} + 2y_{3.} - 4y_{.1})/6. \end{aligned} \tag{12.63}$$

The general estimable function (12.62) provides particular estimable functions, by giving special value to the w's. It can also be used to see if particular functions of interest are estimable (by ascertaining whether or not appropriate values can be chosen for the w's). For example, $\alpha_1 - \alpha_2$ is seen to be estimable, because by putting $w_1 = 1, w_2 = -1$, and $w_3 = 0 = w_4$ in (12.62) it reduces to $\alpha_1 - \alpha_2$. The same values in (12.63) give the BLUE of $\alpha_1 - \alpha_2$. On the other hand, $\mu + \alpha_1$ is not estimable; no values can be chosen for the w's that reduce (12.62) to $\mu + \alpha_1$.

There is, of course, an infinite number of sets of w's that could be used in (12.62), and so there is an infinite number of estimable functions. But each is a function $\mathbf{w}'\mathbf{H}\boldsymbol{\beta}$ for some $\mathbf{w}'$. Since each $\mathbf{w}'$ is arbitrary of order p (the number of elements in $\boldsymbol{\beta}$), we can always find p linearly independent values of $\mathbf{w}'$, represented by $\mathbf{W}$, say. Then $\mathbf{W}^{-1}$ exists, and the corresponding estimable functions are $\mathbf{W}\mathbf{H}\boldsymbol{\beta}$. But $r_{\mathbf{WH}} = r_{\mathbf{H}} = r_{\mathbf{X}}$ (the second equality is true because by Theorem 8.1(ii), the rank of $\mathbf{G}\mathbf{X}'\mathbf{X}$ is equal to the rank of $\mathbf{X}'\mathbf{X}$, which is equal to the rank of $\mathbf{X}$). Therefore, there are no more than $r_{\mathbf{X}}$ linearly independent estimable functions in $\mathbf{W}\mathbf{H}\boldsymbol{\beta}$. Hence, no matter how many particular estimable functions we generate from (12.62) by using a different set of values for the w's, no more than $r_{\mathbf{X}}$ (in this case 4) of them will be linearly independent. Table 12.8 shows such a set of linearly independent estimable functions, and the BLUE of each. These are all familiar results; nevertheless, it is interesting to demonstrate their derivation from the Theorems 12.4 and 12.5.

The estimable functions in Table 12.8 could also be derived from the expected value of y_{ij}. Since from (12.60), the expected value of y_{ij} is

$$E(y_{ij}) = \mu + \alpha_i + \beta_j,$$

any linear combination of the expressions $\mu + \alpha_i + \beta_j$ is estimable. For example,

$$\mu + \alpha_1 + \beta_1 - (\mu + \alpha_2 + \beta_1) = \alpha_1 - \alpha_2$$

is estimable; and, of course, its BLUE is $\widehat{\alpha_1 - \alpha_2} = \alpha_1^0 - \alpha_2^0$, which, from calculating $\boldsymbol{\beta}^0 = \mathbf{G}\mathbf{X}'\mathbf{y}$, is found to be $\bar{y}_{1.} - \bar{y}_{2.}$.

Once again it is appropriate to reiterate what has been said before in these pages. Although these matrix procedures may at first appear cumbersome when applied to the simple, familiar cases just considered, it is the universality of their application to situations not so familiar that commands our attention. With balanced data (the same number of observations in each of the smallest subclassifications of the data), the customary procedures used

TABLE 12.9 Numbers of Plants and Their Yields

	Number of Plants			Yield of Plants		
	Treatment			Treatment		
Variety	1	2	Total	1	2	Total
1	2	1	3	3,7	8	$18 = y_{1.}$
2	1	1	2	34	2	$36 = y_{2.}$
3	1	–	1	12	–	$12 = y_{3.}$
Total	4	2	6	$56 = y_{.1}$	$10 = y_{.2}$	$66 = y_{..}$

for computing parameter estimates are suitable *solely* because the matrix $\mathbf{X}'\mathbf{X}$ has a simple form in such cases. As a result, the matrix methods that have been described lead to these customary procedures. But when data are not balanced, as is so often the case with experimental and field survey data, the expressions for estimates of functions of parameters are not at all standard or well known, and they depend a great deal on the nature of the data available. The matrix methods are then a means of deriving the estimable functions and their estimators. We illustrate this with another example.

12.12.3 Two-Way Classification, No Interactions, Unbalanced Data

Suppose in the previous example there had been two plants of variety 1 receiving treatment 1, and none of variety 3 receiving treatment 2. The numbers of plants and total yields might then be as shown in Table 12.9. The yields are hypothetical and are solely for purposes of illustration; and they have been chosen partly with an eye to easy arithmetic.

The model is essentially the same as before, as are the vectors $\boldsymbol{\beta}$ and $\mathbf{X}'\mathbf{y}$. For the normal equations, $\mathbf{X}'\mathbf{X}$ and a generalized inverse of it are

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 2 & 1 & 4 & 2 \\ 3 & 3 & 0 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 4 & 2 & 1 & 1 & 4 & 0 \\ 2 & 1 & 1 & 0 & 0 & 2 \end{bmatrix}$$

$$\mathbf{G} = \frac{1}{7}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 2 & 4 & -4 & 0 \\ 0 & 2 & 5 & 3 & -3 & 0 \\ 0 & 4 & 3 & 13 & -6 & 0 \\ 0 & -4 & -3 & -6 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{12.64}$$

$\mathbf{H} = \mathbf{G}\mathbf{X}'\mathbf{X}$ turns out to be the same as the previous, and so the estimable functions are the same as in (12.62). But, the BLUE's of these estimable functions, $\mathbf{a}'\boldsymbol{\beta}^0 = \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{y}$ as in

TABLE 12.10 Estimators of the Four Examples of Estimable Functions Shown in Table 12.8

		Estimator	
		Balanced Data Table 12.7	Unbalanced Data Table 12.9
Example	Estimable Function	As in Table 12.8, from Equation (12.63)	Equation (12.65)
1	$\alpha_1 - \alpha_2$	$\bar{y}_{1.} - \bar{y}_{2.}$	$(3y_{1.} - 3y_{2.} + y_{3.} - y_{.1})/7$
2	$\alpha_1 - \alpha_3$	$\bar{y}_{1.} - \bar{y}_{3.}$	$(y_{1.} - y_{2.} - 9y_{3.} + 2y_{.1})/7$
3	$\mu + \bar{\alpha}_. + \bar{\beta}_.$[a]	$\bar{y}_{..}$	$(10y_{1.} + 11y_{2.} + 22y_{3.} - 8y_{.1})/42$
4	$\beta_1 - \beta_2$	$\bar{y}_{1.} - \bar{y}_{2.}$	$-(4y_{1.} + 3y_{2.} + 6y_{3.} - 6y_{.1})/7$

[a] $\bar{\alpha}_. = \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3)$ and $\bar{\beta}_. = \frac{1}{2}(\beta_1 + \beta_2)$.

(12.63), are now

$$\begin{aligned} \boldsymbol{a}'\boldsymbol{\beta}^0 &= \boldsymbol{w}'\boldsymbol{G}(y_{..},\ y_{1.},\ y_{2.},\ y_{3.},\ y_{.1},\ y_{.2})' \qquad (12.65)\\ &= w_1(5y_{1.} + 2_{y2.} + 4y_{3.} - 4y_{.1})/7 + w_2(2y_{1.} + 5y_{2.} + 3y_{3.} - 3y_{.1})/7\\ &\quad + w_3(4y_{1.} + 3y_{2.} + 13y_{3.} - 6y_{.1})/7 - w_4(4y_{1.} + 3y_{2.} + 6y_{3.} - 6y_{.1})/7. \end{aligned}$$

As a function of $y_{1.}, y_{2.}, y_{3.}$, and $y_{.1}$, this expression for $\boldsymbol{a}'\boldsymbol{\beta}^0$ is quite different from its counterpart (12.63) in the balanced data case, even though the expression for the estimable function is the same, namely (12.62). This emphasizes how even a minor difference in the nature of the data available, such as envisaged here, can lead to quite substantial differences in the estimators. Further evidence of this is seen in Table 12.10, which shows the same estimable functions as Table 12.8, with the BLUE's from that table, and with their BLUE's from (12.65). The difference between the two sets of estimators is striking.

As an example of unbalanced data in the two-way classification, that discussed here is about as simple as one can get. The literature abounds with more extensive examples and with description of more general difficulties with this model that are beyond the scope of this book. For further reading, a reference such as Searle et al. (1981), and the sources it uses, provides a starting point.

12.13 THE $R(\cdot|\cdot)$ NOTATION (CONTINUED)

We now extend the ideas implicit in the sum of squares

$$R(\alpha|\mu) = R(\mu,\ \alpha) - R(\mu)$$

introduced in (12.42) and used in the analysis of variance table of Table 12.6. This, it will be recalled, is the sum of squares due to fitting the model $E(y_{ij}) = \mu + \alpha_i$ over and above that due to fitting $E(y_{ij}) = \mu$. The extension is simple. First, in parallel with $R(\alpha|\mu)$, we have

$$R(\beta|\mu) = R(\mu,\ \beta) - R(\mu),$$

TABLE 12.11 Two Partitionings of Total Sum of Squares in the Two-Way Classification Without Interaction

Fitting $\mu + \alpha_i$ Before $\mu + \alpha_i + \beta_j$	Fitting $\mu + \beta_j$ Before $\mu + \alpha_i + \beta_j$
$R(\mu) = N\bar{y}_{..}^2$	$R(\mu) = N\bar{y}_{..}^2$
$R(\alpha\vert\mu) = R(\mu, \alpha) - R(\mu)$	$R(\beta\vert\mu) = R(\mu, \beta) - R(\mu)$
$R(\beta\vert\mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha)$	$R(\alpha\vert\mu, \beta) = R(\mu, \alpha, \beta) - R(\mu, \beta)$
$SS_E = \mathbf{y}'\mathbf{y} - R(\mu, \alpha, \beta)$	$SS_E = \mathbf{y}'\mathbf{y} - R(\mu, \alpha, \beta)$
$SS_T = \mathbf{y}'\mathbf{y}$	$SS_T = \mathbf{y}'\mathbf{y}$

which is the additional sum of squares due to fitting $E(y_{ij}) = \mu + \beta_j$ after fitting $E(y_{ij}) = \mu$. Just like (12.59), for b levels of the β-factor, $j = 1, \ldots, b$,

$$R(\mu,\ \beta) = \sum_{j=1}^{b} n_{.j}\bar{y}_{.j}^2 = \sum_{j=1}^{b} y_{.j}^2/n_{.j}. \tag{12.66}$$

Second, we have

$$R\,(\alpha|\mu,\ \beta) = R\,(\mu,\ \alpha,\ \beta) - R\,(\mu,\ \beta)\,,$$

the sum of squares due to $E(y_{ij}) = \mu + \alpha_i + \beta_j$ after fitting $E(y_{ij}) = \mu + \beta_j$; and in parallel with this there is

$$R\,(\beta|\mu,\ \alpha) = R\,(\mu,\ \alpha,\ \beta) - R\,(\mu,\ \alpha)\,,$$

the sum of squares due to $E(y_{ij}) = \mu + \alpha_i + \beta_j$ after fitting $E(y_{ij}) = \mu + \alpha_i$.

Since, for the model $E(y_{ij}) = \mu + \alpha_i + \beta_j$, the error sum of squares is

$$SS_E = SS_T - R\,(\mu,\ \alpha,\ \beta)$$

with SS_T being the total sum of squares $\boldsymbol{y}'\boldsymbol{y}$, these differences between sums of squares can be summarized as partitionings of SS_T, as shown in Table 12.11. It is similar to, but an extension of, Table 12.6. In calculating these differences, $R(\mu,\ \alpha)$ and $R(\mu,\ \beta)$ come from (12.59) and (12.66), respectively, and $R(\mu,\ \alpha,\ \beta)$ is the reduction in sum of squares for fitting the model

$$E(y_{ij}) = \mu + a_i + \beta_j, \tag{12.67}$$

that is, it is the SS_{reg} for this model. By (12.34) it is therefore

$$R\,(\mu,\ \alpha,\ \beta) = \boldsymbol{\beta}^{\mathbf{0}}\boldsymbol{X}'\boldsymbol{y}, \text{ for (12.67) written as } E(\boldsymbol{y}) = \boldsymbol{X\beta}. \tag{12.68}$$

TABLE 12.12 Degrees of Freedom and Sums of Squares for Analysis of Variance Tables Corresponding to Table 12.11

Fitting $\mu+\alpha_i$ Before $\mu+\alpha_i+\beta_j$		Fitting $\mu+\beta_j$ Before $\mu+\alpha_i+\beta_j$	
DF[a]	SS	DF[a]	SS
$1=1$	$R(\mu)=726$	$1=1$	$R(\mu)=726$
$a-1=2$	$R(\alpha\|\mu)=174$	$b-1=1$	$R(\beta\|\mu)=108$
$b-1=1$	$R(\beta\|\mu,\alpha)=168$	$a-1=2$	$R(\alpha\|\mu,\beta)=234$
$N'=2$	$SS_E=358$	$N'=2$	$SS_E=358$
$N=6$	$SS_T=1426$	$N=6$	$SS_T=1426$

[a] $N'=N-a-b+1=6-2-3+1=2$.

Example 12.10 *For the data of Table 12.9*

$$
\begin{aligned}
R(\mu) &= 6(11)^2 &&= 726\\
R(\mu,\ \alpha) &= 18^2/3+36^2/2+12^2/1 &&= 900\\
R(\mu,\ \beta) &= 56^2/4+10^2/2 &&= 834\\
SS_T &= 3^2+7^2+8^2+34^2+2^2+12^2 &&= 1426.
\end{aligned}
$$

Then, using $\mathbf{G}$ *of (12.64) and*

$$
\mathbf{X}'\mathbf{y}=\begin{bmatrix} y_{\cdot\cdot}\\ y_{1\cdot}\\ y_{2\cdot}\\ y_{3\cdot}\\ y_{\cdot 1}\\ y_{\cdot 2}\end{bmatrix}=\begin{bmatrix}66\\18\\36\\12\\56\\10\end{bmatrix}\quad \textit{gives}\quad \boldsymbol{\beta}^0=\mathbf{G}\mathbf{X}'\mathbf{y}=\begin{bmatrix}0\\-2\\12\\0\\12\\0\end{bmatrix}
$$

so that from (12.68)

$$
\begin{aligned}
R(\mu,\ \alpha,\ \beta) &= 0(66)+(-2)(18)+12(36)+0(12)+12(56)+0(10)\\
&= 1068.
\end{aligned}
$$

The sums of squares of Table 12.11 are shown in Table 12.12. Alongside each is shown the corresponding degrees of freedom used when normality is assumed, and when F-statistics are calculated such as

$$
F(\alpha|\mu)=\frac{R(\alpha|\mu)/(a-1)}{SS_E/N'}=\frac{174/2}{358/2}=\frac{87}{179}
$$

and

$$
F(\beta|\mu,\ \alpha)=\frac{R(\beta|\mu,\ \alpha)/(b-1)}{SS_E/N'}=\frac{168/2}{358/2}=\frac{84}{179}.
$$

TABLE 12.13 Analysis of Variance for Balanced Data

Source of Variation	DF	SS
Mean	1	$N\bar{y}^2_{..}$
Rows (α's)	2	$\sum\sum(\bar{y}_{i\cdot}-\bar{y}_{..})^2$
Columns (β's)	1	$\sum\sum(\bar{y}_{.j}-\bar{y}_{..})^2$
Error	2	$\sum\sum(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$
Total	6	$\sum\sum y_{ij}^2$

Much has been written about F-statistics such as these, and this book is not the place for any extensive discussion. That can be found, for example, in Searle (1971, Chapter 7). It suffices to say, as is shown there (p. 283), that

$$F(\alpha|\mu) \text{ does not test } H_0 : \alpha_i\text{'s all equal.} \tag{12.69}$$

It tests

$$H_0 : \left(\alpha_i + \sum_{j=1}^{b} n_{ij}\beta_j/n_{i\cdot}\right) \text{ equal for all } i,$$

which in the case of this example, is

$$H_0 : \alpha_1 + \frac{2}{3}\beta_1 + \frac{1}{3}\beta_2 = \alpha_2 + \frac{1}{2}\beta_1 + \frac{1}{2}\beta_2 = \alpha_3 + \beta_1. \tag{12.70}$$

In contrast,

$$F(\beta|\mu,\,\alpha) \text{ tests } H_0 : \beta_j\text{'s all equal,}$$

and, correspondingly,

$$F(\alpha|\mu,\,\beta) \text{ tests } H_0 : \alpha_i\text{'s all equal,}$$

The existence of distinctly different partitionings of the total sum of squares as shown in Table 12.11 and illustrated in Table 12.12 arises solely from there being unequal numbers of observations in the subclasses of the data. When there is the same number in every subclass, as in Table 12.7, both partitionings reduce, algebraically, to the familiar forms shown in Table 12.13 (wherein all double summations are over $i = 1, \ldots, a = 3$ *and* $j = 1, \ldots, b = 2$*). For example, with balanced data, both* $R(\alpha|\mu)$ *and* $R(\alpha|\mu,\,\beta)$ *simplify to be* $\sum\sum(\bar{y}_{i\cdot}-\bar{y}_{..})^2$*. This simplification is due* solely *to the form of the various* $\boldsymbol{X}$*-matrices involved when calculating each reduction in sum of squares as* $SS_{reg} = \boldsymbol{\beta}^{0\prime}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{XGx}'\boldsymbol{y}$ *of (12.34) when data are balanced.*

The matrix development presented here is always appropriate for unbalanced data (even in the presence of covariates—see Searle (1971), Section 8.1). It also applies to balanced data (in that, for example, both parts of Table 12.12 simplify to the same thing, namely Table 12.13). There is therefore some convenience in thinking of the analysis of balanced data as being a special case of unbalanced data and not vice versa .

The whole subject of analyzing unbalanced data has been touched on here in terms of just the simplest of examples. But this is an effective introduction to the general ideas involved. They can readily be extended to more complex situations, but to do so here would be beyond the scope of this book. The current trend is to eliminate many of the complexities of these over-parameterized (more parameters to be estimated than there are means to estimate them from) models by returning to the basic model as, for example, $E(y_{ijk}) = \mu_{ij}$ *[see, e.g., Searle (1987) and Hocking (1996)]. This is called the* μ_{ij}-model, *or* cell means model—*but it is another subject altogether.*

12.14 REPARAMETERIZATION TO A FULL-RANK MODEL

The objective here is to provide a certain reparameterization that introduces a full-rank model for the response vector $\boldsymbol{y}$. Consider the model in (12.1). Applying the spectral decomposition theorem (Theorem 6.2) to the symmetric matrix $\boldsymbol{X}'\boldsymbol{X}$, we get

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{P}\mathrm{diag}(\boldsymbol{\Lambda}, \boldsymbol{0})\boldsymbol{P}',$$

where, $\boldsymbol{\Lambda}$ is an $r \times r$ diagonal matrix of r nonzero eigenvalues, and $\boldsymbol{0}$ is a diagonal matrix of zero eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. The columns of $\boldsymbol{P}$ are corresponding orthonormal eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$. We can partition $\boldsymbol{P}$ as $[\boldsymbol{P}_1 \;\; \boldsymbol{P}_2]$ where the columns of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ correspond to $\boldsymbol{\Lambda}$ and $\boldsymbol{0}$, respectively. The matrix $\boldsymbol{X}'\boldsymbol{X}$ can then be written as $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{P}_1\boldsymbol{\Lambda}\boldsymbol{P}_1'$. Note that $\boldsymbol{P}_1$ is of order $p \times r$ and rank r. We also have $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{P}_2 = \boldsymbol{0}$ since the columns of $\boldsymbol{P}_2$ are eigenvectors corresponding to the zero eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. This implies that $\boldsymbol{X}\boldsymbol{P}_2 = \boldsymbol{0}$. Model (12.1) can then be expressed as

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{P}\boldsymbol{P}'\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
&= [\boldsymbol{X}\boldsymbol{P}_1 \;\; \boldsymbol{0}] \begin{bmatrix} \boldsymbol{P}_1'\boldsymbol{\beta} \\ \boldsymbol{P}_2'\boldsymbol{\beta} \end{bmatrix} + \boldsymbol{\epsilon} \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{P}_1\boldsymbol{P}_1'\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
\boldsymbol{y} &= \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon},
\end{aligned} \tag{12.71}$$

where $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{P}_1$ and $\boldsymbol{\delta} = \boldsymbol{P}_1'\boldsymbol{\beta}$. The matrix $\boldsymbol{Z}$, which is of order $n \times r$, is of full column rank due to the fact that

$$\begin{aligned}
r(\boldsymbol{Z}) &= r(\boldsymbol{Z}'\boldsymbol{Z}) \\
&= r(\boldsymbol{P}_1'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{P}_1) \\
&= r(\boldsymbol{P}_1'\boldsymbol{P}_1\boldsymbol{\Lambda}\boldsymbol{P}_1'\boldsymbol{P}_1) \\
&= r(\boldsymbol{\Lambda}) \\
&= r,
\end{aligned}$$

since $\boldsymbol{P}_1'\boldsymbol{P}_1 = \boldsymbol{I}_r$. The model $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon}$ is then a full-rank reparameterization of model (12.1) with $\boldsymbol{\delta} = \boldsymbol{P}_1'\boldsymbol{\beta}$ being a linear function of $\boldsymbol{\beta}$.

REFERENCES

Christensen, R. (1996). *Analysis of Variance, Design, and Regression*. Chapman & Hall/CRC, Boca Raton, FL.

Federer, W. T. (1955). *Experimental Design*. Macmillan, New York.

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

Graybill, F. A. (1976). *Theory and Applications of the Linear Model*. Duxbury Press, North Scituate, MA.

Hinkelman, K. and Kempthorne, O. (1994). *Design and Analysis of Experiments*. John Wiley & Sons, New York.

Hocking, R. R. (1996). *Methods and Applications of Linear Models*. John Wiley & Sons, New York.

Jorgensen, B. (1993). *The Theory of Linear Models*. Chapman & Hall, New York.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. John Wiley & Sons, New York.

Khuri, A. I., Mathew, T., and Sinha, B. K. (1998). *Statistical Tests for Mixed Linear Models*. John Wiley & Sons, New York.

Khuri, A. I. (2010). *Linear Model Methodology*. Chapman & Hall/CRC, Boca Raton, FL.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley & Sons, New York.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models*. Springer-Verlag, New York.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons, New York.

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

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. John Wiley & Sons, New York.

Searle, S. R., Speed, F. M., and Henderson, H. V. (1981). Some computational and model equivalences in analysis of variance of unequal-subclass-numbers data. *Am. Stat.*, 35, 16–33.

Seber, G. A. F. (1977). *Linear Regression Analysis*. John Wiley & Sons, New York.

Sengupta, D. and Jammalamadaka, S. R. (2003). *Linear Models*. World Scientific, Singapore.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*. McGraw-Hill, New York.

EXERCISES

12.1 Deduce from model (12.71) that the number of linearly independent estimable functions of $\boldsymbol{\beta}$ in model (12.1) is equal to the rank of $\boldsymbol{X}$.

12.2 Consider the matrix $\boldsymbol{P}_1'$ used in Section (12.14). Show that the rows of $\boldsymbol{P}_1'$ form an orthonormal basis for the row space of $\boldsymbol{X}$.

12.3 Consider model (12.1) and let $\boldsymbol{Q}'\boldsymbol{\beta}$ be estimable, where $\boldsymbol{Q}'$ is of order $s \times p$ and rank s. Let $\boldsymbol{\beta}^{\mathbf{0}}$ be a solution to the normal equations. Show that $\boldsymbol{Q}'\boldsymbol{\beta}^{\mathbf{0}}$ is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^{-}$.

12.4 Give details of the proof of the Gauss–Markov Theorem (Theorem 12.3).

12.5
(a) Show that any linear function of $\boldsymbol{\beta}$ in model (12.1) is estimable if and only if it can be expressed as $\boldsymbol{l}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$ for some vector $\boldsymbol{l}'$. This expression gives the general form of all estimable linear functions of $\boldsymbol{\beta}$.

(b) Deduce that the coefficients of the distinct elements of $\boldsymbol{l}'$ in $\boldsymbol{l}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$ form a basis for the space of all estimable linear functions of $\boldsymbol{\beta}$.

12.6 Consider the following data set:

	B		
A	1	2	3
1	18	12	24
2	5	7	9
3	3	6	15
4	6	3	18

for which the appropriate model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$.

(a) Find a solution to the normal equations.

(b) Give a test statistic for the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3$. Show first that the linear functions $\beta_1 - \beta_2$ and $\beta_1 - \beta_3$ are estimable.

(c) Find the power of the test in part (b) at the 5% level of significance given that the error variance $\sigma^2 = 1$ and $\beta_1 - \beta_2 = 1$, $\beta_1 - \beta_3 = 3$.

(d) Show that $\mu + \frac{1}{4}\sum_{i=1}^{4} \alpha_i + \beta_1$ is estimable and find its BLUE.

Note: The following SAS statements can be used to obtain the solutions to parts (a), (b), and (c) of this exercise:

```
DATA ONE;
INPUT A B YY@@;
CARDS;
1 1 18 1 2 12 1 3 24
2 1 5 2 2 7 2 3 9
3 1 3 3 2 6 3 3 15
4 1 6 4 2 3 4 3 18
PROC PRINT;
PROC GLM;
CLASS A B;
MODEL YY=A B/E SOLUTION;
RUN;
PROC IML;
X={1 1 0 0 0 1 0 0,
1 1 0 0 0 0 1 0,
1 1 0 0 0 0 0 1,
1 0 1 0 0 1 0 0,
1 0 1 0 0 0 1 0,
```

```
1 0 1 0 0 0 0 1,
1 0 0 1 0 1 0 0,
1 0 0 1 0 0 1 0,
1 0 0 1 0 0 0 1,
1 0 0 0 1 1 0 0,
1 0 0 0 1 0 1 0,
1 0 0 0 1 0 0 1};
PRINT X;
Y={18 12 24 5 7 9 3 6 15 6 3 18}‘;
PRINT Y;
XPX=X‘*X;
GXPX=GINV(XPX);
MSE=Y‘*(I(12) - X*GXPX*X‘)*Y/6;
PRINT MSE;
BETA=GXPX*X‘*Y;
A={0 0 0 0 0 1 -1 0 0 0 0 0 0 1 0 -1 };
PRINT A;
M={0,0 };
MA={1,3 };
F=(A*BETA - m)‘*(INV(A*GXPX*A‘))*(A*BETA - M)/(2*MSE);
PRINT F;
THETA=(MA - M)‘*(INV(A*GXPX*A‘))*(MA - M);
POWER=1 - PROBF(5.14,2,6,THETA);
PRINT POWER;
```

Note that the use of the two at-symbols,@@, at the end of the INPUT statement allows the data to be entered continuously in several lines. If these symbols are not used, then the data must be entered on separate lines where each line contains only one value for each of A, B, YY. A detailed explanation of the above SAS statements will be given in Chapter 15 of Part III of this book.

12.7 Consider again the data set of Exercise 6. Assume that the cells (2,1), (2,2), and (3,2) are empty, that is, have no observations. Show that their corresponding cell means are still estimable.

12.8 Consider again Exercise 7. Show that $\phi = \sum_{i=1}^{4} \lambda_i \alpha_i$ is estimable if $\sum_{i=1}^{4} \lambda_i = 0$, that is, ϕ is a contrast in the α_i's. Find its BLUE.

12.9 Consider once more the data set of Exercise 6. Compute the following sums of squares:

(a) $R(\mu)$.

(b) $R(\alpha|\mu)$.

(c) $R(\mu, \alpha)$.

(d) $R(\alpha|\mu, \beta)$.

(e) $R(\beta|\mu, \alpha)$.

12.10 Consider the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

where $\epsilon_{ijk} \sim N(0, \sigma^2)$. The corresponding data set used with this model is

	B			
A	1	2	3	4
1	12	.	24	30
2	.	.	6	11
3	21	15	16	.

(a) What is the number of basic linearly independent estimable linear functions for this data set? Specify a set of such functions that spans the space of all estimable linear functions for this model.

(b) Is $\mu_{12} = \mu + \alpha_1 + \beta_2$ estimable? Why or why not?

(c) Show that $\frac{1}{4}\sum_{j=1}^{4}(\mu + \alpha_i + \beta_j)$ is estimable for $i = 1, 2, 3$ and find its BLUE.

Note: The following SAS statements (see Chapter 15) can be used to find the solutions to this exercise:

```
DATA;
INPUT A B Y@@;
CARDS;
1 1 12 1 2 . 1 3 24 1 4 30
2 1 . 2 2 . 2 3 6 2 4 11
3 1 21 3 2 15 3 3 16 3 4 .
PROC PRINT;
PROC GLM;
CLASS A B;
MODEL Y = A B/E SOLUTION;
RUN;
```

An explanation of these SAS statements will be given in Chapter 15 of Part III of this book.

13

Analysis of Balanced Linear Models Using Direct Products of Matrices

This chapter demonstrates the usefulness of direct products of matrices in the study of balanced linear models. It is shown that the derivation of the general properties of such models can be greatly facilitated by using direct products. This includes the determination of the distribution of sums of squares in a balanced model with random and fixed effects under the assumptions of independence and normality of the random effects in the model. Furthermore, by expressing the model in terms of direct products of identity matrices and vectors of ones, several matrix identities can be easily derived. It can therefore be stated that the use of direct products with balanced models provides a well-structured foundation upon which a clear understanding of the properties of such models can be developed.

Before we begin our study of balanced linear models, a number of definitions need to be introduced in addition to the development of a certain notation for the general treatment of such models. The material presented here is based on Chapter 8 in Khuri (2010).

Definition 13.1 *A data set associated with a response variable y is said to be balanced if the range of any one subscript of y does not depend on the values of the other subscripts of y. A linear model used in the analysis of such a data set is called a balanced linear model.*

Definition 13.2 (Crossed and Nested Factors) *Factors A and B are said to be crossed if every level of one factor is used in combination with every level of the other factor. If, however, the levels of one factor, for example B, depend on the values of the levels of another factor, for example A, then B is said to be nested within A. In this case, the levels of B used in combination with a level of A are different entities from those used with other levels of A.*

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

For example, tomato plants were grown in a greenhouse under treatments consisting of combinations of soil type (factor A) and fertilizer type (factor B). Three replications of the yield y of tomatoes were obtained per treatment combination. This is an example of a factorial experiment with two crossed factors and equal numbers of observations in the various cells. In a different experimental situation, an experiment involving four drugs was conducted to study each drug's effect on the heart rate y of human subjects. At the start of the study, 10 male subjects were randomly assigned to each drug (the subjects were different for different drugs). This is obviously a nested experiment with subjects (factor B) being nested within drugs (factor A).

Notation If i and j are subscripts associated with A and B, respectively, and if A and B are crossed, then i and j are said to be *crossed subscripts* and this is denoted by writing $(i)(j)$. If, however, factor B is nested within factor A, then subscript j is said to be *nested* within subscript i. This is denoted by writing $(i : j)$ where a colon separates the two subscripts with the subscript appearing after the colon being the nested subscript. In this case, it should be noted that since j depends on i, the identification of a level of B also requires the identification of the level of A nesting it. Thus whenever j appears as a subscript, subscript i must also appear.

Definition 13.3 (Population Structure) *For a given experiment, the identification of the crossed and nested relationships that exist among the subscripts defines the so-called population structure.*

For example, [(i)(j)]:k is the population structure for a factorial experiment with two crossed factors and several replications per treatment combination. Here, subscript k serves to identify the observation obtained under a treatment combination. Also, (i:j:k) is a population structure for a nested experiment with several replications made on the response y as a result of using level j of B which is nested within level i of A. The description of a population structure is needed in order to set up a complete model for the experiment in addition to the specification of the corresponding analysis of variance (ANOVA).

Definition 13.4 (Partial Mean) *A partial mean of a response y is the average value of y obtained over the range of a particular set of subscripts. A partial mean is denoted by the same symbol as the one used for the response y with subscripts, not including the subscripts that have been averaged out.*

For example, for an experiment with two crossed factors and several replications on the response y per treatment combination, y_i, y_j, y_{ij} are partial means which, under the traditional notation, correspond to $\bar{y}_{i..}$, $\bar{y}_{.j.}$, and $\bar{y}_{ij.}$, respectively.

Definition 13.5 (Admissible Mean) *A partial mean is said to be admissible if whenever a nested subscript appears, all the subscripts that nest it appear also.*

For example, for a nested experiment with two factors, A and B, where B is nested within A, the means with subscripts i and ij are admissible, but a mean with only subscript j is not admissible.

TABLE 13.1 Population Structures and Admissible Means

Population Structure	Admissible Means
$[(i)(j)] : k$	$y, y_{(i)}, y_{(j)}, y_{(ij)}, y_{ij(k)}$
$(i : j : k : l)$	$y, y_{(i)}, y_{i(j)}, y_{ij(k)}, y_{ijk(l)}$
$[(i : j)(k)] : l$	$y, y_{(i)}, y_{(k)}, y_{i(j)}, y_{(ik)}, y_{i(jk)}, y_{ijk(l)}$

Definition 13.6 (Rightmost-Bracket Subscripts) *The rightmost bracket of an admissible mean is a subset of subscripts (of the mean) that nest no other subscripts of that mean. These subscripts are identified by using a pair of parentheses that contain them. The remaining subscripts, if any, of the admissible mean are called nonrightmost-bracket subscripts and are placed before the rightmost bracket. If an admissible mean has no subscripts, as is case with the overall mean of the response, then the admissible mean is denoted by writing just y.*

For example, the admissible means corresponding to the three population structures, $[(i)(j)] : k$, $(i : j : k : l)$, and $[(i : j)(k)] : l$ are given in Table 13.1.

The same notation concerning admissible means can be extended to their corresponding effects in a given model. As will be seen later, the rightmost bracket for an effect plays an important role in determining the degrees of freedom, sums of squares, and expected mean squares for that effect in the corresponding ANOVA table.

Definition 13.7 (Component) *A component corresponding to an admissible mean is a linear combination of admissible means that are yielded by the mean being considered when some, all, or none of its rightmost-bracket subscripts are omitted in all possible ways. If an odd number of subscripts is omitted, the resulting admissible mean is assigned a negative sign, otherwise, if an even number of subscripts is omitted, the mean is assigned a positive sign. (the number zero is considered even).*

For example, the components corresponding to the admissible means for each of the three population structures in Table 13.1 are displayed in Table 13.2.

It is important here to note that for any given population structure, the sum of all components corresponding to its admissible means is identical to the response y. Since these components correspond in a one-to-one fashion with the effects in the associated model, such an identity gives rise to a complete model for the response under consideration. It can therefore be concluded that knowledge of the population structure in an experimental situation is very important in determining the complete model for the response, in addition to identifying the various effects that make up the model.

13.1 GENERAL NOTATION FOR BALANCED LINEAR MODELS

We now present a general notation for the study of balanced linear models. This notation will be instrumental in the derivation of properties associated with these models. The same notation can also be found in Khuri (2010) and a similar one in Khuri (1982).

Let $\omega = \{k_1, k_2, \ldots, k_s\}$ be a complete set of subscripts that identify a typical response y, where $k_j = 1, 2, \ldots, a_j$ $(j = 1, 2, \ldots, s)$. It can be noted that the corresponding data set is

TABLE 13.2 Admissible Means and Corresponding Components

Population Structure	Admissible Mean	Component
$[(i)(j)]:k$	y	y
	$y_{(i)}$	$y_{(i)} - y$
	$y_{(j)}$	$y_{(j)} - y$
	$y_{(ij)}$	$y_{(ij)} - y_{(i)} - y_{(j)} + y$
	$y_{ij(k)}$	$y_{ij(k)} - y_{(ij)}$
$(i:j:k:l)$	y	y
	$y_{(i)}$	$y_{(i)} - y$
	$y_{i(j)}$	$y_{i(j)} - y_{(i)}$
	$y_{ij(k)}$	$y_{ij(k)} - y_{i(j)}$
	$y_{ijk(l)}$	$y_{ijk(l)} - y_{ij(k)}$
$[(i:j)(k)]:l$	y	y
	$y_{(i)}$	$y_{(i)} - y$
	$y_{(k)}$	$y_{(k)} - y$
	$y_{i(j)}$	$y_{i(j)} - y_{(i)}$
	$y_{(ik)}$	$y_{(ik)} - y_{(i)} - y_{(k)} + y$
	$y_{i(jk)}$	$y_{i(jk)} - y_{i(j)} - y_{(ik)} + y_{(i)}$
	$y_{ijk(l)}$	$y_{ijk(l)} - y_{i(jk)}$

balanced since the range of any one subscript does not depend on the values of the other subscripts. Let N denote the total number of observations in the data set, that is, $N = \prod_{j=1}^{s} a_j$.

Suppose that for a given population structure, the number of admissible means is $q + 2$. The ith admissible mean is denoted by $y_{\omega_i(\bar{\omega}_i)}$, where $\bar{\omega}_i$ is the set of rightmost-bracket subscripts and ω_i is the set of nonrightmost-bracket subscripts ($i = 0, 1, 2, \ldots, q + 1$). For $i = 0$, both ω_i and $\bar{\omega}_i$ are empty. For some other admissible means, ω_i may also be empty. The set of of all subscripts that belong to both ω_i and $\bar{\omega}_i$ is denoted by η_i. The complement of η_i with respect to ω is denoted by η_i^c ($i = 0, 1, \ldots, q + 1$). Note that $\eta_i = \omega$ when $i = q + 1$. The ith component corresponding to the ith admissible mean is denoted by $C_{\omega_i(\bar{\omega}_i)}(y)$ ($i = 0, 1, \ldots, q + 1$).

In general, any balanced linear model can be expressed as

$$y_\omega = \sum_{i=0}^{q+1} f_{\omega_i(\bar{\omega}_i)}, \tag{13.1}$$

where $f_{\omega_i(\bar{\omega}_i)}$ denotes the ith effect in the model. For $i = 0, f_{\omega_i(\bar{\omega}_i)}$ is the grand mean μ and for $i = q + 1, f_{\omega_i(\bar{\omega}_i)}$ is the experimental error term. Furthermore, the ith component, $C_{\omega_i(\bar{\omega}_i)}(y)$, can in general be written as

$$C_{\omega_i(\bar{\omega}_i)}(y) = \sum_{j=0}^{q+1} \lambda_{ij} y_{\omega_j(\bar{\omega}_j)}, \quad i = 0, 1, \ldots, q + 1, \tag{13.2}$$

where $\lambda_{ij} = -1, 0, 1$. The values -1 and 1 are obtained whenever an odd number or an even number of subscripts are omitted from $\bar{\omega}_i$, respectively. The value $\lambda_{ij} = 0$ applies to those admissible means that are not obtained by deleting subscripts from $\bar{\omega}_i$.

Using direct products, the vector form of model (13.1) can be expressed as

$$\boldsymbol{y} = \sum_{i=0}^{q+1} \boldsymbol{U}_i \boldsymbol{\beta}_i, \tag{13.3}$$

where $\boldsymbol{y}$ denotes the vector of N observations, $\boldsymbol{\beta}_i$ is a vector consisting of the elements of $f_{\omega_i(\bar{\omega}_i)}$ $(i = 0, 1, \ldots, q+1)$, and the matrix $\boldsymbol{U}_i$ is a direct (Kronecker) product of matrices of the form

$$\boldsymbol{U}_i = \otimes_{j=1}^{s} \boldsymbol{L}_{ij}, \quad i = 0, 1, \ldots, q+1, \tag{13.4}$$

where for each i $(= 0, 1, \ldots, q+1)$, the matrices $\boldsymbol{L}_{ij}$ $(j = 1, 2, \ldots, s)$ correspond in a one-to-one fashion to the elements of $\omega = (k_1, k_2, \ldots, k_s)$ such that

$$\boldsymbol{L}_{ij} = \begin{cases} \boldsymbol{I}_{a_j}, & k_j \in \eta_i \\ \boldsymbol{1}_{a_j}, & k_j \in \eta_i^c \end{cases}; \; i = 0, 1, \ldots, q+1; \, j = 1, 2, \ldots, s, \tag{13.5}$$

where, η_i is the set of subscripts associated with the ith effect, η_i^c is its complement with respect to ω $(i = 0, 1, \ldots, q+1)$, and $\boldsymbol{I}_{a_j}$ and $\boldsymbol{1}_{a_j}$ are, respectively, the identity matrix of order $a_j \times a_j$ and the vector of ones of order $a_j \times 1$ $(j = 1, 2, \ldots, s)$. In other words, if a subscript of $\omega = (k_1, k_2, \ldots, k_s)$ belongs to η_i, the corresponding $\boldsymbol{L}_{ij}$ matrix in (13.4) is equal to $\boldsymbol{I}_{a_j}$, otherwise, it is equal to $\boldsymbol{1}_{a_j}$ if it does not belong to η_i $(j = 1, 2, \ldots, s)$.

Using (13.4) we get

$$\boldsymbol{U}_i'\boldsymbol{U}_i = \otimes_{j=1}^{s} \boldsymbol{L}_{ij}'\boldsymbol{L}_{ij}.$$

It is easy to see that $\boldsymbol{L}_{ij}'\boldsymbol{L}_{ij}$ is equal to $\boldsymbol{I}_{a_j}$ if k_j belongs to η_i, otherwise, it is equal to a_j if k_j does not belong to η_i. Hence, $\boldsymbol{U}_i'\boldsymbol{U}_i$ must be equal to a scalar multiple of the identity matrix which is of order $c_i \times c_i$, where

$$c_i = \prod_{k_j \in \eta_i} a_j, \quad i = 0, 1, \ldots, q+1. \tag{13.6}$$

In the event $\eta_i = \emptyset$, where $\emptyset$ denotes the empty set, then $c_i = 1$. The scalar multiple of the identity matrix is denoted by b_i, where

$$b_i = \prod_{k_j \in \eta_i^c} a_j, \quad i = 0, 1, \ldots, q+1. \tag{13.7}$$

If $\eta_i^c = \emptyset$, then $b_i = 1$. We therefore have

$$\boldsymbol{U}_i'\boldsymbol{U}_i = b_i \boldsymbol{I}_{c_i}, \quad i = 0, 1, \ldots, q+1. \tag{13.8}$$

Note that c_i represents the number of columns of $\boldsymbol{U}_i$ while b_i is equal to the number of ones in a column of $\boldsymbol{U}_i$ $(i = 0, 1, \ldots, q+1)$. Thus, $b_i c_i = N$ for $i = 0, 1, \ldots, q+1$.

Example 13.1 *Consider the two-way crossed classification model*

$$y_{ijk} = \mu + \alpha_{(i)} + \beta_{(j)} + (\alpha\beta)_{(ij)} + \epsilon_{ij(k)}, \tag{13.9}$$

where $i = 1, 2, 3$, $j = 1, 2, 3, 4$, $k = 1, 2, 3$. *The corresponding population structure is [(i)(j)]:k. Note that the model has been written using the new notation where the rightmost bracket is identified in each one of its effects. The vector form of the model is*

$$\boldsymbol{y} = \boldsymbol{U}_0\mu + \boldsymbol{U}_1\boldsymbol{\alpha} + \boldsymbol{U}_2\boldsymbol{\beta} + \boldsymbol{U}_3(\boldsymbol{\alpha\beta}) + \boldsymbol{U}_4\boldsymbol{\epsilon}, \tag{13.10}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $(\boldsymbol{\alpha\beta})$ *contain the elements of* $\boldsymbol{\alpha}_{(i)}$, $\boldsymbol{\beta}_{(j)}$, $(\boldsymbol{\alpha\beta})_{(ij)}$, *respectively. The matrices* $\boldsymbol{U}_0$, $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, $\boldsymbol{U}_3$, $\boldsymbol{U}_4$ *are given by*

$$\begin{aligned}
\boldsymbol{U}_0 &= \boldsymbol{1}_3 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_1 &= \boldsymbol{I}_3 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_2 &= \boldsymbol{1}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_3 &= \boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_4 &= \boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_3.
\end{aligned}$$

Example 13.2 *Consider the population structure (i:j:k:l) for which the model is the three-fold nested model,*

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{i(j)} + \gamma_{ij(k)} + \epsilon_{ijk(l)}, \tag{13.11}$$

where $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$, $k = 1, 2, 3$, $l = 1, 2, 3$. *The corresponding vector form is*

$$\boldsymbol{y} = \boldsymbol{U}_0\mu + \boldsymbol{U}_1\boldsymbol{\alpha} + \boldsymbol{U}_2\boldsymbol{\beta} + \boldsymbol{U}_3\boldsymbol{\gamma} + \boldsymbol{U}_4\boldsymbol{\epsilon}, \tag{13.12}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ *contain the elements of* $\boldsymbol{\alpha}_{(i)}$, $\boldsymbol{\beta}_{i(j)}$, $\boldsymbol{\gamma}_{ij(k)}$, *respectively, and the* $\boldsymbol{U}_i$ *matrices are given by*

$$\begin{aligned}
\boldsymbol{U}_0 &= \boldsymbol{1}_4 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_1 &= \boldsymbol{I}_4 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_2 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{1}_3 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_3 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_3 \otimes \boldsymbol{1}_3\\
\boldsymbol{U}_4 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_3 \otimes \boldsymbol{I}_3.
\end{aligned}$$

Example 13.3 *Consider the third population structure given in Table 13.2, namely,* $[(i : j)(k)] : l$. *The corresponding model is*

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{i(j)} + \gamma_{(k)} + (\alpha\gamma)_{(ik)} + (\beta\gamma)_{i(jk)} + \epsilon_{ijk(l)},$$

where $i = 1, 2, 3, 4$; $j = 1, 2, 3, 4$; $k = 1, 2, 3, 4$; $l = 1, 2, 3$. *The vector form of the model is*

$$\boldsymbol{y} = \boldsymbol{U}_0\mu + \boldsymbol{U}_1\boldsymbol{\alpha} + \boldsymbol{U}_2\boldsymbol{\beta} + \boldsymbol{U}_3\boldsymbol{\gamma} + \boldsymbol{U}_4(\boldsymbol{\alpha\gamma}) + \boldsymbol{U}_5(\boldsymbol{\beta\gamma}) + \boldsymbol{U}_6\boldsymbol{\epsilon},$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $(\boldsymbol{\alpha\gamma})$, and $(\boldsymbol{\beta\gamma})$ contain the elements of $\alpha_{(i)}$, $\beta_{i(j)}$, $\gamma_{(k)}$, $(\alpha\gamma)_{(ik)}$, $(\beta\gamma)_{i(jk)}$, respectively, and

$$\begin{aligned}
\boldsymbol{U}_0 &= \mathbf{1}_4 \otimes \mathbf{1}_4 \otimes \mathbf{1}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_1 &= \boldsymbol{I}_4 \otimes \mathbf{1}_4 \otimes \mathbf{1}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_2 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \mathbf{1}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_3 &= \mathbf{1}_4 \otimes \mathbf{1}_4 \otimes \boldsymbol{I}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_4 &= \boldsymbol{I}_4 \otimes \mathbf{1}_4 \otimes \boldsymbol{I}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_5 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \mathbf{1}_3\\
\boldsymbol{U}_6 &= \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_3.
\end{aligned}$$

13.2 PROPERTIES ASSOCIATED WITH BALANCED LINEAR MODELS

Balanced linear models have very interesting properties which will be presented in this section. These properties set balanced models apart from other types of models in the sense that they provide general rules that facilitate the analysis of data generated by such models. The properties are presented here as lemmas, some of which are proved while others are merely stated. The interested reader is referred to Zyskind (1962) and Khuri (1982, 2010) for more details.

Lemma 13.1 *The sum of values of any component is zero when the summation is taken over any subset of subscripts in its rightmost bracket. In other words, if $C_{\omega_i(\bar{\omega}_i)}(y)$ is the ith component, and if $\bar{\omega}_i \neq \emptyset$, then*

$$\sum_{\tau_i \subset \bar{\omega}_i} C_{\omega_i(\bar{\omega}_i)}(y) = 0,$$

where τ_i denotes a subset of subscripts in $\bar{\omega}_i$.

Definition 13.8 *The sum of squares for the ith effect in model (13.1) is written as a quadratic form $\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}$ $(i = 0, 1, \ldots, q+1)$ defined as*

$$\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y} = \sum_{\omega}[C_{\omega_i(\bar{\omega}_i)}(y)]^2, \quad i = 0, 1, \ldots, q+1. \tag{13.13}$$

For example, for the three-fold nested model in Example 13.2, the sums of squares for the various effects in the model are (see Table 13.1)

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{P}_0\boldsymbol{y} &= \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{3}\sum_{l=1}^{3} y^2\\
&= 144\, y^2\\
\boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y} &= \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=3}^{3}\sum_{l=1}^{3} [y_{(i)} - y]^2\\
&= 36\sum_{i=1}^{4}[y_{(i)} - y]^2
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{y}'\boldsymbol{P}_2\boldsymbol{y} &= \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{3}\sum_{l=1}^{3}[y_{i(j)} - y_{(i)}]^2 \\
&= 9\sum_{i=1}^{4}\sum_{j=1}^{4}[y_{i(j)} - y_{(i)}]^2 \\
\boldsymbol{y}'\boldsymbol{P}_3\boldsymbol{y} &= \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{3}\sum_{l=1}^{3}[y_{ij(k)} - y_{i(j)}]^2 \\
&= 3\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{3}[y_{ij(k)} - y_{i(j)}]^2 \\
\boldsymbol{y}'\boldsymbol{P}_4\boldsymbol{y} &= \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{3}\sum_{l=1}^{3}[y_{ijk(l)} - y_{ij(k)}]^2.
\end{aligned}
$$

The following lemma expresses a sum of squares in an ANOVA table as a certain linear combination of sums of squares of admissible means.

Lemma 13.2 *The sum of squares for the ith effect in model (13.1) can be expressed as*

$$
\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y} = \sum_{j=0}^{q+1}\lambda_{ij}\sum_{\omega} y^2_{\omega_j(\bar{\omega}_j)}, \quad i = 0, 1, \ldots, q+1, \tag{13.14}
$$

where λ_{ij} is the same as the coefficient of the admissible mean $y_{\omega_j(\bar{\omega}_j)}$ in the expression given by (13.2) which defines the ith component ($i, j = 0, 1, \ldots, q+1$). In other words, the ith sum of squares is a linear combination of the sums of squares (over the set of all subscripts, ω) of the admissible means that make up the ith component.

For example, for the model in Example 13.3 with the the population structure $[(i : j)(k)]{:}l$ its sums of squares can be expressed as

$$
\begin{aligned}
\boldsymbol{y}'\boldsymbol{P}_0\boldsymbol{y} &= 192\,y^2 \\
\boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y} &= 48\sum_{i=1}^{4} y^2_{(i)} - 192\,y^2 \\
\boldsymbol{Y}'\boldsymbol{P}_2\boldsymbol{y} &= 12\sum_{i=1}^{4}\sum_{j=1}^{4} y^2_{i(j)} - 48\sum_{i=1}^{4} y^2_{(i)} \\
\boldsymbol{y}'\boldsymbol{P}_3\boldsymbol{y} &= 48\sum_{k=1}^{4} y^2_{(k)} - 192\,y^2 \\
\boldsymbol{y}'\boldsymbol{P}_4\boldsymbol{y} &= 12\sum_{i=1}^{4}\sum_{k=1}^{4} y^2_{(ik)} - 48\sum_{i=1}^{4} y^2_{(i)} - 48\sum_{k=1}^{4} y^2_{(k)} + 192\,y^2
\end{aligned}
$$

$$\boldsymbol{y}'\boldsymbol{P}_5\boldsymbol{y} = 3\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} y_{i(jk)}^2 - 12\sum_{i=1}^{4}\sum_{j=1}^{4} y_{i(j)}^2 - 12\sum_{i=1}^{4}\sum_{k=1}^{4} y_{(ik)}^2 + 48\sum_{i=1}^{4} y_{(i)}^2$$

$$\boldsymbol{y}'\boldsymbol{P}_6\boldsymbol{y} = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4}\sum_{l=1}^{3} y_{ijk(l)}^2 - 3\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} y_{i(jk)}^2.$$

The $\boldsymbol{P}_i$ matrices have certain properties which are described in the following lemma.

Lemma 13.3 *The matrix $\boldsymbol{P}_i$ in formula (13.13) has the following properties:*

(a) $\boldsymbol{P}_i$ *is idempotent* $(i = 0, 1, \ldots, q+1)$
(b) $\boldsymbol{P}_i\boldsymbol{P}_j = \boldsymbol{0}$ *for* $i \neq j$ $(i, j = 0, 1, \ldots, q+1)$
(c) $\sum_{i=0}^{q+1} \boldsymbol{P}_i = \boldsymbol{I}_N$.

Lemma 13.4 *The matrix $\boldsymbol{P}_i$ can be expressed as*

$$\boldsymbol{P}_i = \sum_{j=0}^{q+1} \lambda_{ij}\frac{\boldsymbol{A}_j}{b_j}, \quad i = 0, 1, \ldots, q+1, \tag{13.15}$$

where $\boldsymbol{A}_j = \boldsymbol{U}_j\boldsymbol{U}_j'$ $(j = 0, 1, \ldots, q+1)$, the λ_{ij}'s are the same coefficients as in (13.2), and b_j is the number defined in (13.7).

This representation of $\boldsymbol{P}_i$ is analogous to the one for the ith component in (13.2) in terms of admissible means. In the present case, the admissible means are replaced by the $\frac{\boldsymbol{A}_j}{b_j}$ matrices.

Proof. Let $\boldsymbol{y}_i$ denote the vector consisting of all the values of the ith admissible mean, $y_{\omega_i(\bar{\omega}_i)}$ $(i = 0, 1, \ldots, q+1)$. Then, it is easy to see that

$$\boldsymbol{y}_i = \frac{1}{b_i}\boldsymbol{U}_i'\boldsymbol{y}, \quad i = 0, 1, \ldots, q+1. \tag{13.16}$$

Using formulas (13.14) and (13.16) we get

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y} &= \sum_{j=0}^{q+1} \lambda_{ij} \sum_{\omega} y_{\omega_j(\bar{\omega}_j)}^2 \\
&= \sum_{j=0}^{q+1} \lambda_{ij}\, b_j \sum_{\eta_j} y_{\omega_j(\bar{\omega}_j)}^2 \\
&= \sum_{j=0}^{q+1} \lambda_{ij}\, b_j \left(\frac{\boldsymbol{y}'\boldsymbol{A}_j\boldsymbol{y}}{b_j^2}\right) \\
&= \boldsymbol{y}'\left(\sum_{j=0}^{q+1} \lambda_{ij}\frac{\boldsymbol{A}_j}{b_j}\right)\boldsymbol{y}.
\end{aligned} \tag{13.17}$$

From (13.17) we conclude (13.15). ∎

Lemma 13.5 *Consider formula (13.15). Then,*

$$\frac{\boldsymbol{A}_j}{b_j} = \sum_{\eta_i \subset \eta_j} \boldsymbol{P}_i, \;\; j = 0, 1, \ldots, q+1. \tag{13.18}$$

The summation in (13.18) extends over all those i subscripts for which $\eta_i \subset \eta_j$ for a given j $(i, j = 0, 1, \ldots, q+1)$.

Proof. Formula (13.15) is derived from formula (13.2) by replacing $C_{\omega_i(\bar{\omega}_i)}(y)$ with $\boldsymbol{P}_i$ and $y_{\omega_j(\bar{\omega}_j)}$ with $\frac{\boldsymbol{A}_j}{b_j}$. It can be seen that the sum of all components whose sets of subscripts are contained inside η_j is equal to $y_{\omega_j(\bar{\omega}_j)}$, that is,

$$y_{\omega_j(\bar{\omega}_j)} = \sum_{\eta_i \subset \eta_j} C_{\omega_i(\bar{\omega}_i)}(y), \;\; j = 0, 1, \ldots, q+1.$$

Replacing $y_{\omega_j(\bar{\omega}_j)}$ by $\frac{\boldsymbol{A}_j}{b_j}$ and $C_{\omega_i(\bar{\omega}_i)}(y)$ by $\boldsymbol{P}_i$, we get (13.18). ∎

The result of the next lemma will be very useful in the analysis of balanced linear models.

Lemma 13.6

$$\boldsymbol{A}_j \boldsymbol{P}_i = \kappa_{ij} \boldsymbol{P}_i, \;\; i, j = 0, 1, \ldots, q+1, \tag{13.19}$$

where $\boldsymbol{A}_j = \boldsymbol{U}_j \boldsymbol{U}_j'$, and

$$\kappa_{ij} = \begin{cases} b_j, & \eta_i \subset \eta_j \\ 0, & \eta_i \not\subset \eta_j \end{cases} ; \;\; i, j = 0, 1, \ldots, q+1. \tag{13.20}$$

Proof. Multiplying the two sides of (13.18) on the right by $\boldsymbol{P}_i$, we get

$$\boldsymbol{A}_j \boldsymbol{P}_i = b_j \left(\sum_{\eta_\ell \subset \eta_j} \boldsymbol{P}_\ell \right) \boldsymbol{P}_i, \;\; i, j = 0, 1, \ldots, q+1.$$

If $\eta_i \subset \eta_j$, then

$$\left(\sum_{\eta_\ell \subset \eta_j} \boldsymbol{P}_\ell \right) \boldsymbol{P}_i = \boldsymbol{P}_i,$$

since $\boldsymbol{P}_i$ is idempotent, and $\boldsymbol{P}_i \boldsymbol{P}_\ell = \boldsymbol{0}$ if $i \neq \ell$ by Lemma 13.3. If, however, $\eta_i \not\subset \eta_j$, then

$$\left(\sum_{\eta_\ell \subset \eta_j} \boldsymbol{P}_\ell \right) \boldsymbol{P}_i = \boldsymbol{0},$$

since the $\boldsymbol{P}_i$'s are orthogonal by Lemma 13.3. It follows that $\boldsymbol{A}_j \boldsymbol{P}_i = \kappa_{ij} \boldsymbol{P}_i$, where κ_{ij} is defined by (13.20). ∎

Lemma 13.7 *Let m_i be the rank of $\boldsymbol{P}_i$ in (13.13). Then, m_i is the same as the number of degrees of freedom for the ith effect in model (13.1), and is equal to*

$$m_i = \left[\prod_{k_j \in \omega_i} a_j\right]\left[\prod_{k_j \in \bar{\omega}_i} (a_j - 1)\right], \quad i = 0, 1, \ldots, q+1, \tag{13.21}$$

where $\bar{\omega}_i$ and ω_i are, respectively, the rightmost- and nonrightmost-bracket subscripts for the ith effect ($i = 0, 1, \ldots, q+1$).

Example 13.4 *Consider the two-way model used in Example 13.1. Let $\boldsymbol{P}_i$ be the matrix associated with the ith sum of squares for this model ($i = 0, 1, \ldots, 4$). Using formula (13.15) and the fact that $a_1 = 3$, $a_2 = 4$, $a_3 = 3$, we get*

$$\begin{aligned}
\boldsymbol{P}_0 &= \frac{1}{36}\boldsymbol{A}_0 \\
\boldsymbol{P}_1 &= \frac{1}{12}\boldsymbol{A}_1 - \frac{1}{36}\boldsymbol{A}_0 \\
\boldsymbol{P}_2 &= \frac{1}{9}\boldsymbol{A}_2 - \frac{1}{36}\boldsymbol{A}_0 \\
\boldsymbol{P}_3 &= \frac{1}{3}\boldsymbol{A}_3 - \frac{1}{12}\boldsymbol{A}_1 - \frac{1}{9}\boldsymbol{A}_2 + \frac{1}{36}\boldsymbol{A}_0 \\
\boldsymbol{P}_4 &= \boldsymbol{A}_4 - \frac{1}{3}\boldsymbol{A}_3,
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{A}_0 &= \boldsymbol{U}_0\boldsymbol{U}_0' = \boldsymbol{J}_3 \otimes \boldsymbol{J}_4 \otimes \boldsymbol{J}_3 \\
\boldsymbol{A}_1 &= \boldsymbol{U}_1\boldsymbol{U}_1' = \boldsymbol{I}_3 \otimes \boldsymbol{J}_4 \otimes \boldsymbol{J}_3 \\
\boldsymbol{A}_2 &= \boldsymbol{U}_2\boldsymbol{U}_2' = \boldsymbol{J}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{J}_3 \\
\boldsymbol{A}_3 &= \boldsymbol{U}_3\boldsymbol{U}_3' = \boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{J}_3 \\
\boldsymbol{A}_4 &= \boldsymbol{U}_4\boldsymbol{U}_4' = \boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_3
\end{aligned}$$

Solving for $\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3, \boldsymbol{A}_4$ in terms of $\boldsymbol{P}_0, \boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$, we get

$$\begin{aligned}
\boldsymbol{A}_0 &= 36\boldsymbol{P}_0 \\
\boldsymbol{A}_1 &= 12(\boldsymbol{P}_0 + \boldsymbol{P}_1) \\
\boldsymbol{A}_2 &= 9(\boldsymbol{P}_0 + \boldsymbol{P}_2) \\
\boldsymbol{A}_3 &= 3(\boldsymbol{P}_0 + \boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3) \\
\boldsymbol{A}_4 &= \boldsymbol{P}_0 + \boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3 + \boldsymbol{P}_4.
\end{aligned}$$

The same results can be obtained from applying formula (13.18). Using formula (13.19), it can be verified that, for example,

$$\begin{gathered}
\boldsymbol{A}_1\boldsymbol{P}_1 = 12\boldsymbol{P}_1, \quad \boldsymbol{A}_1\boldsymbol{P}_2 = \boldsymbol{0}, \quad \boldsymbol{A}_2\boldsymbol{P}_1 = \boldsymbol{0} \\
\boldsymbol{A}_2\boldsymbol{P}_2 = 9\boldsymbol{P}_2, \quad \boldsymbol{A}_3\boldsymbol{P}_1 = 3\boldsymbol{P}_1, \quad \boldsymbol{A}_3\boldsymbol{P}_2 = 3\boldsymbol{P}_2 \\
\boldsymbol{A}_4\boldsymbol{P}_3 = \boldsymbol{P}_3, \quad \boldsymbol{A}_4\boldsymbol{P}_2 = \boldsymbol{P}_2, \quad \boldsymbol{A}_3\boldsymbol{P}_4 = \boldsymbol{0}.
\end{gathered}$$

13.3 ANALYSIS OF BALANCED LINEAR MODELS

Let us now consider the nature of the effects in the general balanced linear model (13.1). If all the effects in the model are randomly distributed, except for the term corresponding to $i = 0$, then the model is called a *random-effects model* (or just a *random model*). If the model contains only fixed unknown parameters, except for the experimental error term, then it is called a *fixed-effects model* (or just a *fixed model*). If, however, the model contains fixed effects (besides the term corresponding to $i = 0$) and at least one random effect (besides the experimental error), then it is called a *mixed-effects model* (or just a *mixed model*).

In this section, we discuss the distribution of the sums of squares in the ANOVA table corresponding to a general balanced model, as in (13.1), when the model is of the fixed-effects, random-effects, or mixed-effects type. We consider the effects corresponding to $i = 0, 1, \ldots, q - p$ to be fixed and those corresponding to $i = q - p + 1, q - p + 2, \ldots, q + 1$ to be random, where p is a non-negative integer not exceeding q. If $p = 0$, then the model is of the fixed-effects type; if $p = q$, then the model is of the random-effects type, otherwise, if $0 < p < q$, then it is of the mixed-effects type. The vector form of the balanced linear model shown in (13.3) can then be written as

$$\boldsymbol{y} = \boldsymbol{Xg} + \boldsymbol{Zh}, \tag{13.22}$$

where $\boldsymbol{Xg} = \sum_{i=0}^{q-p} \boldsymbol{U}_i\boldsymbol{\beta}_i$ is the fixed portion of the model and $\boldsymbol{Zh} = \sum_{i=q-p+1}^{q+1} \boldsymbol{U}_i\boldsymbol{\beta}_i$ is its random portion. It is assumed that $\boldsymbol{\beta}_{q-p+1}, \boldsymbol{\beta}_{q-p+2}, \ldots, \boldsymbol{\beta}_{q+1}$ are mutually independent and normally distributed such that

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{0}, \sigma_i^2\boldsymbol{I}_{c_i}), \quad i = q - p + 1, q - p + 2, \ldots, q + 1, \tag{13.23}$$

where c_i is the number of columns of $\boldsymbol{U}_i$ [see (13.6)]. Under these assumptions, $E(\boldsymbol{y}) = \boldsymbol{Xg}$, and the variance-covariance matrix, $\boldsymbol{\Sigma}$, of $\boldsymbol{y}$ is

$$\boldsymbol{\Sigma} = \sum_{i=q-p+1}^{q+1} \sigma_i^2\boldsymbol{A}_i, \tag{13.24}$$

where $\boldsymbol{A}_i = \boldsymbol{U}_i\boldsymbol{U}_i^{\prime}$ $(i = q - p + 1, q - p + 2, \ldots, q + 1)$.

13.3.1 Distributional Properties of Sums of Squares

The properties outlined in Section 13.2 can now be put to use to derive the distribution of the sums of squares of the effects in a balanced mixed model. The following theorem is considered the cornerstone of the analysis of balanced linear models. Details of its proof can be found in Khuri (2010, Section 8.4.1).

Theorem 13.1 *Let $\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}$ be the sum of squares of the ith effect for model (13.22), $i = 0, 1, \ldots, q + 1$. Then, under the assumptions described earlier concerning the model's random effects,*

(a) $\boldsymbol{y}'\boldsymbol{P}_0\boldsymbol{y}, \boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y}, \ldots, \boldsymbol{y}'\boldsymbol{P}_{q+1}\boldsymbol{y}$ *are mutually independent.*

(b)
$$\frac{\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}}{\rho_i} \sim \chi^2_{m_i}(\lambda_i), \quad i = 0, 1, \ldots, q + 1, \tag{13.25}$$

where m_i is the number of degrees of freedom for the ith effect (which is the same as the rank of $\boldsymbol{P}_i$), λ_i is the noncentrality parameter which is equal to

$$\lambda_i = \frac{\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_i\boldsymbol{X}\boldsymbol{g}}{\rho_i}, \quad i = 0, 1, \ldots, q+1, \tag{13.26}$$

and

$$\rho_i = \sum_{j \in W_i} b_j \sigma_j^2, \quad i = 0, 1, \ldots, q+1, \tag{13.27}$$

where b_j is defined in formula (13.7) and W_i is the set

$$W_i = \{j : q-p+1 \leq j \leq q+1 \mid \eta_i \subset \eta_j\}, \quad i = 0, 1, \ldots, q+1, \tag{13.28}$$

where, if we recall, η_i is the set of subscripts that identifies the ith effect (Section 13.1).

(c)
$$E(\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}) = \boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_i\boldsymbol{X}\boldsymbol{g} + m_i\,\rho_i, \quad i = 0, 1, \ldots, q+1. \tag{13.29}$$

(d) *The noncentrality parameter λ_i in part (b) is equal to zero if the ith effect is random, that is, for $i = q-p+1,\ q-p+2, \ldots, q+1$. Thus, for such an effect, $\frac{1}{\rho_i}\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}$ has the central chi-squared distribution with m_i degrees of freedom and $E(\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}) = m_i\rho_i$.*

Note that the expression for ρ_i in (13.27) contains only those σ^2's whose sets of subscripts (that is, η_j) contain the set of subscripts that identifies the *i*th effect (that is, η_i). Each one of these σ^2's has a coefficient equal to the product of the ranges of those subscripts that do not appear in the designation of that particular σ^2. The error variance, denoted by σ_ϵ^2, appears in all the ρ_i's with a coefficient equal to one. We also note that the expression for $\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_i\boldsymbol{X}\boldsymbol{g}$ in (13.29) is the same as $\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}$ with $\boldsymbol{y}$ replaced by $\boldsymbol{X}\boldsymbol{g}$, which is the mean of $\boldsymbol{y}$. Therefore, to compute $\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_i\boldsymbol{X}\boldsymbol{g}$ we can use the formula for $\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y}$ in (13.13), namely,

$$\boldsymbol{y}'\boldsymbol{P}_i\boldsymbol{y} = \sum_{\omega} [C_{\omega_i(\bar{\omega}_i)}(y)]^2, \quad i = 0, 1, \ldots, q+1,$$

and then replace the admissible means in $C_{\omega_i(\bar{\omega}_i)}(y)$ with their expected values.

Example 13.5 *Consider the balanced mixed two-fold nested model*

$$y_{ijk} = \mu + \alpha_{(i)} + \beta_{i(j)} + \epsilon_{ij(k)}, \tag{13.30}$$

where $\alpha_{(i)}$ $(i = 1, 2, 3)$ is a fixed unknown parameter and $\beta_{i(j)}$ and $\epsilon_{ij(k)}$ are randomly distributed as $N(0, \sigma^2_{\beta(\alpha)})$ and $N(0, \sigma_\epsilon^2)$, respectively $(j = 1, 2, 3, 4;\ k = 1, 2, 3)$. All random effects are independent. The expected mean squares can be obtained by applying formula (13.29). These mean squares are denoted by MS_A, $MS_{B(A)}$, MS_E, and correspond to $\alpha_{(i)}$, $\beta_{i(j)}$, $\epsilon_{ij(k)}$, respectively. If we number the effects in model (13.30) as 0, 1, 2, 3, then

$$\begin{aligned} E(MS_A) &= \frac{1}{2}\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g} + \rho_1 \\ E(MS_{B(A)}) &= \rho_2 \\ E(MS_E) &= \rho_3, \end{aligned}$$

TABLE 13.3 ANOVA Table for Model (13.30)

Source	DF	SS	MS	E(MS)
A	2	SS_A	MS_A	$6\sum_{i=1}^{3}\alpha_{(i)}^2 + 3\sigma_{\beta(\alpha)}^2 + \sigma_\epsilon^2$
$B(A)$	9	$SS_{B(A)}$	$MS_{B(A)}$	$3\sigma_{\beta(\alpha)}^2 + \sigma_\epsilon^2$
Error	24	SS_E	MS_E	σ_ϵ^2

since $m_1 = 2$, *where* $\rho_1 = 3\,\sigma_{\beta(\alpha)}^2 + \sigma_\epsilon^2$, $\rho_2 = 3\,\sigma_{\beta(\alpha)}^2 + \sigma_\epsilon^2$, $\rho_3 = \sigma_\epsilon^2$. *As was mentioned earlier,* $\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}$ *can be obtained by using the formula for* SS_A, *namely,*

$$SS_A = \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{3}[y_{(i)} - y]^2,$$

and then replacing $y_{(i)}$ *and* y *with their expected values, namely,*

$$E(y_{(i)}) = \mu + \alpha_{(i)}$$
$$E(y) = \mu,$$

we get

$$\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g} = 12\sum_{i=1}^{3}\alpha_{(i)}^2.$$

The corresponding ANOVA table is shown in Table 13.3.

According to Theorem 13.1, we have

$$\frac{SS_A}{\rho_1} \sim \chi_2^2(\lambda_1),$$

where $\lambda_1 = \dfrac{\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}}{\rho_1} = \dfrac{12\sum_{i=1}^{3}\alpha_{(i)}^2}{3\,\sigma_{\beta(\alpha)}^2+\sigma_\epsilon^2}$. *Furthermore,*

$$\frac{SS_{B(A)}}{\rho_2} = \frac{SS_{B(A)}}{3\,\sigma_{\beta(\alpha)}^2 + \sigma_\epsilon^2} \sim \chi_9^2$$
$$\frac{SS_E}{\rho_3} = \frac{SS_E}{\sigma_\epsilon^2} \sim \chi_{24}^2.$$

It can therefore be seen that a test statistic for testing $H_0 : \alpha_{(i)} = 0$ *is given by the F-ratio,* $\frac{MS_A}{MS_{B(A)}}$, *which under* H_0 *has the F-distribution with 2 and 9 degrees of freedom. Furthermore, to test the hypothesis* $H_0 : \sigma_{\beta(\alpha)}^2 = 0$, *we can use the F-ratio,* $\frac{MS_{B(A)}}{MS_E}$, *which under* H_0 *has the F-distribution 9 and 24 degrees of freedom.*

13.3.2 Estimates of Estimable Linear Functions of the Fixed Effects

The purpose of this section is to show how to derive estimates of estimable linear functions of $\boldsymbol{g}$, the vector of fixed effects in model (13.22). A basis for all such functions is described in Theorem 13.2. Details of its proof can be found in Khuri (2010, Section 8.4.2).

Theorem 13.2 *Let $\boldsymbol{P}_i$ be the matrix associated with the sum of squares for the ith effect ($i = 0, 1, \ldots, q+1$) as shown in (13.13). Then,*

(a) *For the ith fixed effect in model (13.22), rank($\boldsymbol{P}_i\boldsymbol{X}$) $= m_i$, where $m_i =$ rank($\boldsymbol{P}_i$), $i = 0, 1, \ldots, q-p$.*
(b) *rank($\boldsymbol{X}$) $= \sum_{i=0}^{q-p} m_i$.*
(c) *$\boldsymbol{P}_0\boldsymbol{X}\boldsymbol{g}, \boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}, \ldots, \boldsymbol{P}_{q-p}\boldsymbol{X}\boldsymbol{g}$ are linearly independent.*
(d) *Any estimable linear function of $\boldsymbol{g}$ can be written as the sum of linear functions of $\boldsymbol{P}_0\boldsymbol{X}\boldsymbol{g}, \boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}, \ldots, \boldsymbol{P}_{q-p}\boldsymbol{X}\boldsymbol{g}$.*

The next theorem shows how to obtain the BLUE of a vector of estimable linear functions of the fixed-effects vector, $\boldsymbol{g}$.

Theorem 13.3 *Let $\boldsymbol{Q}'\boldsymbol{g}$ be a vector of estimable linear functions of $\boldsymbol{g}$. Then, the best linear unbiased estimator of $\boldsymbol{Q}'\boldsymbol{g}$ is given by $\boldsymbol{Q}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$.*

The proof of this theorem requires the following lemmas:

Lemma 13.8 *$\boldsymbol{P}_i\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{P}_i$ for $i = 0, 1, \ldots, q+1$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix in (13.24).*

This follows from the fact that $\boldsymbol{\Sigma}\boldsymbol{P}_i = \rho_i\boldsymbol{P}_i$ ($i = 0, 1, \ldots, q+1$). The proof of this assertion is as follows:

Using formulas (13.19) and (13.24), we have

$$\begin{aligned}\boldsymbol{\Sigma}\boldsymbol{P}_i &= \left(\sum_{j=q-p+1}^{q+1} \sigma_j^2 \boldsymbol{A}_j\right)\boldsymbol{P}_i \\ &= \left(\sum_{j=q-p+1}^{q+1} \kappa_{ij}\sigma_j^2\right)\boldsymbol{P}_i, \quad i = 0, 1, \ldots, q+1.\end{aligned}$$

From formula (13.20) we get

$$\sum_{j=q-p+1}^{q+1} \kappa_{ij}\sigma_j^2 = \sum_{j\in W_i} b_j\sigma_j^2,$$

where $\boldsymbol{W}_i$ is the set defined in (13.28). Hence, by using formula (13.27), we get

$$\boldsymbol{\Sigma}\boldsymbol{P}_i = \rho_i\boldsymbol{P}_i, \quad i = 0, 1, \ldots, q+1. \tag{13.31}$$

Taking the transpose of both side of (13.31) we conclude that $\boldsymbol{P}_i\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{P}_i$.

Lemma 13.9 *$\boldsymbol{A}_i\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{A}_i$ for $i = 0, 1, \ldots, q+1$, where $\boldsymbol{A}_i = \boldsymbol{U}_i\boldsymbol{U}_i'$ and $\boldsymbol{\Sigma}$ is the same as in Lemma 13.8.*

Proof. From Lemma 13.5, we get

$$\boldsymbol{A}_i = b_i \sum_{\eta_j \subset \eta_i} \boldsymbol{P}_j, \; i = 0, 1, \ldots, q+1.$$

Hence,

$$\begin{aligned}
\boldsymbol{A}_i\boldsymbol{\Sigma} &= b_i \left(\sum_{\eta_j \subset \eta_i} \boldsymbol{P}_j \right) \boldsymbol{\Sigma} \\
&= b_i \sum_{\eta_j \subset \eta_i} (\boldsymbol{\Sigma}\boldsymbol{P}_j), \text{ by Lemma 13.8,} \\
&= \boldsymbol{\Sigma} \left(b_i \sum_{\eta_j \subset \eta_i} \boldsymbol{P}_j \right) \\
&= \boldsymbol{\Sigma}\boldsymbol{A}_i.
\end{aligned}$$

■

Lemma 13.10 *There exists a nonsingular matrix $\boldsymbol{F}$ such that $\boldsymbol{\Sigma X} = \boldsymbol{XF}$.*

Proof. $\boldsymbol{\Sigma X} = \boldsymbol{\Sigma}[\boldsymbol{U}_0 : \boldsymbol{U}_1 : \ldots : \boldsymbol{U}_{q-p}]$. From Lemma 13.9, $\boldsymbol{A}_i\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{A}_i$, $i = 0, 1, \ldots, q+1$. Then,

$$\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{U}_i\boldsymbol{U}_i', \; i = 0, 1, \ldots, q+1.$$

Multiplying both sides on the right by $\boldsymbol{U}_i$ and recalling from (13.8) that $\boldsymbol{U}_i'\boldsymbol{U}_i = b_i\boldsymbol{I}_{c_i}$ ($i = 0, 1, \ldots, q+1$), we get

$$\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{\Sigma}\boldsymbol{U}_i = b_i\boldsymbol{\Sigma}\boldsymbol{U}_i, \; i = 0, 1, \ldots, q+1.$$

Hence,

$$\begin{aligned}
\boldsymbol{\Sigma}\boldsymbol{U}_i &= \frac{1}{b_i}\boldsymbol{U}_i\boldsymbol{U}_i'\boldsymbol{\Sigma}\boldsymbol{U}_i \\
&= \boldsymbol{U}_i\boldsymbol{G}_i, \; i = 0, 1, \ldots, q+1,
\end{aligned}$$

where $\boldsymbol{G}_i = \frac{1}{b_i}\boldsymbol{U}_i'\boldsymbol{\Sigma}\boldsymbol{U}_i$. Note that $\boldsymbol{G}_i$ is a $c_i \times c_i$ matrix of rank c_i and is therefore nonsingular. It follows that

$$\begin{aligned}
\boldsymbol{\Sigma X} &= [\boldsymbol{\Sigma U}_0 : \boldsymbol{\Sigma U}_1 : \ldots : \boldsymbol{\Sigma U}_{q-p}] \\
&= [\boldsymbol{U}_0\boldsymbol{G}_0 : \boldsymbol{U}_1\boldsymbol{G}_1 : \ldots : \boldsymbol{U}_{q-p}\boldsymbol{G}_{q-p}] \\
&= [\boldsymbol{U}_0 : \boldsymbol{U}_1 : \ldots : \boldsymbol{U}_{q-p}] \bigoplus_{i=0}^{q-p} \boldsymbol{G}_i \\
&= \boldsymbol{XF},
\end{aligned}$$

where $\boldsymbol{F} = \bigoplus_{i=0}^{q-p} \boldsymbol{G}_i$, which is nonsingular. ■

Proof of Theorem 13.3 The best linear unbiased estimator (BLUE) of $\boldsymbol{Q}'\boldsymbol{g}$ is $\boldsymbol{Q}'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. By Lemma 13.10, $\boldsymbol{\Sigma X} = \boldsymbol{XF}$ for some nonsingular matrix $\boldsymbol{F}$. Hence,

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{\Sigma X F}^{-1} \\ \boldsymbol{X}'\boldsymbol{\Sigma}^{-1} &= (\boldsymbol{F}^{-1})'\boldsymbol{X}'. \end{aligned}$$

Consequently,

$$(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = [(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{X}]^{-}(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{y}. \tag{13.32}$$

Note that $[(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{X}]^{-}$ can be chosen equal to $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{F}'$ since

$$(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{F}'(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{X} = (\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{X}.$$

From (13.32) we conclude that

$$\begin{aligned} (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} &= (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{F}'(\boldsymbol{F}^{-1})'\boldsymbol{X}'\boldsymbol{y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}. \end{aligned}$$

This indicates that the BLUE of $\boldsymbol{Q}'\boldsymbol{g}$ is $\boldsymbol{Q}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}$.

From Theorem 13.3 we conclude that the BLUE of $\boldsymbol{Q}'\boldsymbol{g}$ can be obtained without requiring knowledge of $\boldsymbol{\Sigma}$, which is usually unknown. Moreover, this estimator is identical to the ordinary least-squares estimator of $\boldsymbol{Q}'\boldsymbol{g}$.

Theorem 13.4 *The variance-covariance matrix of the best linear unbiased estimator of $\boldsymbol{Q}'\boldsymbol{g}$ in Theorem 13.3 is given by*

$$var[\boldsymbol{Q}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}] = \boldsymbol{T}'\left(\sum_{i=0}^{q-p}\rho_i\boldsymbol{P}_i\right)\boldsymbol{T},$$

where $\boldsymbol{T}$ is such that $\boldsymbol{Q}' = \boldsymbol{T}'\boldsymbol{X}$, and ρ_i is given in (13.27) ($i = 0, 1, \ldots, q-p$).

Details of the proof of this theorem can be found in Khuri (2010, Section 8.4.2).

Example 13.6 *Consider the balanced mixed two-way crossed classification model*

$$y_{ijk} = \mu + \alpha_{(i)} + \beta_{(j)} + (\alpha\beta)_{(ij)} + \epsilon_{ij(k)}, \tag{13.33}$$

where $\alpha_{(i)}$ is a fixed unknown parameter and $\beta_{(j)}$, $(\alpha\beta)_{(ij)}$, and $\epsilon_{ij(k)}$ are randomly distributed as $N(0, \sigma^2_{\beta})$, $N(0, \sigma^2_{\alpha\beta})$, and $N(0, \sigma^2_{\epsilon})$, respectively. All random effects are independent. The corresponding data set for this model is given in Table 13.4. The corresponding ANOVA table is given in Table 13.5. The expected mean squares in the ANOVA table were obtained by applying formula (13.29). If we number the effects in model (13.33) as 0, 1, 2, 3, and 4, respectively, then

$$\begin{aligned} E(MS_A) &= \boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g} + \rho_1 \\ E(MS_B) &= \rho_2 \\ E(MS_{AB}) &= \rho_3 \\ E(MS_E) &= \rho_4, \end{aligned}$$

TABLE 13.4 Data Set for Model (13.33)

	B		
A	1	2	3
1	72	75	92
	60	71	101
	62	72	91
2	101	110	116
	96	108	110
	109	101	106

where $\rho_1 = 3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$, $\rho_2 = 6\,\sigma^2_{\beta} + 3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$, $\rho_3 = 3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$, $\rho_4 = \sigma^2_{\epsilon}$, *as can be seen from applying formula (13.27). Furthermore,* $\boldsymbol{X} = [\boldsymbol{U}_0 : \boldsymbol{U}_1]$, *where* $\boldsymbol{U}_0 = \boldsymbol{1}_2 \otimes \boldsymbol{1}_3 \otimes \boldsymbol{1}_3$, $\boldsymbol{U}_1 = \boldsymbol{I}_2 \otimes \boldsymbol{1}_3 \otimes \boldsymbol{1}_3$, $\boldsymbol{g} = [\mu, \alpha_{(1)}, \alpha_{(2)}]'$, *and from (13.15),* $\boldsymbol{P}_1$ *is given by*

$$\boldsymbol{P}_1 = \frac{1}{9}\boldsymbol{I}_2 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_3 - \frac{1}{18}\boldsymbol{J}_2 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_3. \tag{13.34}$$

A more direct way to compute $\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}$ *is to apply the technique mentioned at the end of Theorem 13.1. In this example, this amounts to replacing* $\boldsymbol{y}$ *in* $SS_A = \boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y}$ *with* $\boldsymbol{X}\boldsymbol{g}$, *which is the mean of* $\boldsymbol{y}$. *The formula for* SS_A *is*

$$SS_A = 9\sum_{i=1}^{2}[y_{(i)} - y]^2. \tag{13.35}$$

We now replace $y_{(i)}$ *and* y *with their expected values, respectively. From model (13.33), these expected values are*

$$E[y_{(i)}] = \mu + \alpha_{(i)}$$
$$E(y) = \mu.$$

Hence,

$$\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g} = 9\sum_{i=1}^{2}\alpha^2_{(i)}.$$

This more direct method to computing $\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_1\boldsymbol{X}\boldsymbol{g}$ *can be applied in general to any fixed effect in a balanced mixed linear model: If the ith effect is fixed, then*

$$\boldsymbol{g}'\boldsymbol{X}'\boldsymbol{P}_i\boldsymbol{X}\boldsymbol{g} = \sum_{\omega} f^2_{\omega_i(\bar{\omega}_i)}, \tag{13.36}$$

TABLE 13.5 ANOVA Table for Model (13.33)

Source	DF	SS	MS	E(MS)	*F*	*p*-Value
A	1	3784.50000	3784.50000	$9\sum_{i=1}^{2}\alpha^2_{(i)} + 3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$	19.39	0.0479
B	2	1170.3333	585.1667	$6\,\sigma^2_{\beta} + 3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$	3.0000	0.2501
$A * B$	2	390.3333	195.1667	$3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}$	7.03	0.0096
Error	12	333.3333	27.7778	σ^2_{ϵ}		

where from Section 13.1, ω is the complete set of subscripts that identify the response y in the model under consideration, and $f_{\omega_i(\bar{\omega}_i)}$ is the ith effect in the model [see model (13.1)].

Estimation of Estimable Linear Function of the Fixed Effects: *Suppose that it is of interest to estimate $\alpha_{(1)} - \alpha_{(2)}$, which is estimable since $\alpha_{(1)} - \alpha_{(2)} = (\mu + \alpha_{(1)}) - (\mu + \alpha_{(2)})$ and both $\mu + \alpha_{(1)}$ and $\mu + \alpha_{(2)}$ are estimable. Using Theorem 13.3, the BLUE of $\alpha_{(1)} - \alpha_{(2)}$ is $\hat{\alpha}_{(1)} - \hat{\alpha}_{(2)}$, where $\hat{\alpha}_{(1)}$ and $\hat{\alpha}_{(2)}$ are obtained from*

$$\begin{aligned} \hat{\boldsymbol{g}} &= [\hat{\mu}, \hat{\alpha}_{(1)}, \hat{\alpha}_{(2)}]' \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \\ &= (61.2222, 16.1111, 45.1111)'. \end{aligned}$$

Thus, $\hat{\alpha}_{(1)} - \hat{\alpha}_{(2)} = 16.1111 - 45.1111 = -29$. Furthermore, since

$$y_{(i)} = \mu + \alpha_{(i)} + \frac{1}{3}\sum_{j=1}^{3}\beta_{(j)} + \frac{1}{3}\sum_{j=1}^{3}(\alpha\beta)_{(ij)} + \frac{1}{9}\sum_{k=1}^{3}\sum_{j=1}^{3}\epsilon_{ij(k)},$$

then

$$var[y_{(1)} - y_{(2)}] = \frac{2}{9}(3\,\sigma^2_{\alpha\beta} + \sigma^2_{\epsilon}). \tag{13.37}$$

Hence, the 95% confidence interval on $\alpha_{(1)} - \alpha_{(2)}$ is given by

$$\begin{aligned} \hat{\alpha}_{(1)} - \hat{\alpha}_{(2)} \pm \left(\frac{2}{9}MS_{AB}\right)^{1/2} t_{0.025,2} &= -29 \pm \left[\frac{2}{9}(195.1667)\right]^{1/2} (4.303) \\ &= -29 \pm 28.3379, \end{aligned}$$

which yields the interval (−57.3379, −0.6621). This indicates that a significant difference can be detected between the two means of factor A, which coincides with the outcome of the F-test for A from Table 13.5 (p-value = 0.0479).

Example 13.7 *Consider the mixed three-way crossed classification model*

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{(j)} + \gamma_{(k)} + (\alpha\beta)_{(ij)} + (\alpha\gamma)_{(ik)} + (\beta\gamma)_{(jk)} + (\alpha\beta\gamma)_{(ijk)} + \epsilon_{ijk(l)}, \tag{13.38}$$

where $\alpha_{(i)}$ and $\beta_{(j)}$ are fixed and $\gamma_{(k)}$ is random. The population structure is [(i)(j)(k)]:l ($i = 1, 2, \ldots, a; j = 1, 2, \ldots, b; k = 1, 2, \ldots, c; l = 1, 2, \ldots, n$). The corresponding variance components are σ^2_{γ}, $\sigma^2_{\alpha\gamma}$, $\sigma^2_{\beta\gamma}$, $\sigma^2_{\alpha\beta\gamma}$, and σ^2_{ϵ}. The expected mean squares and appropriate hypotheses are given in Table 13.6.

The quadratic forms for the fixed effects of $\alpha_{(i)}$, $\beta_{(j)}$, and $(\alpha\beta)_{(ij)}$ were obtained using the short-cut method mentioned earlier. Using the expected mean squares in Table 13.6, it can be seen that the tests for the various hypotheses are shown in Table 13.7 which displays the F-test statistic for each hypothesis along with its degrees of freedom. A large value exceeding the corresponding α-critical value is significant at the α-level.

Note that no single mean square exists in Table 13.6 that can be used as an "error" term for testing the hypothesis $H_0 : \sigma^2_{\gamma} = 0$. In this case, it is possible to create an "artificial"

TABLE 13.6 Expected Mean Squares and Null Hypotheses

E(MS)	Null Hypothesis
$E(MS_A) = \frac{nbc}{a-1}\sum_{i=1}^{a}\alpha_{(i)}^2 + bn\sigma_{\alpha\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \alpha_{(i)} = 0$
$E(MS_B) = \frac{nac}{b-1}\sum_{j=1}^{b}\beta_j^2 + an\sigma_{\beta\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \beta_{(j)} = 0$
$E(MS_C) = nab\sigma_\gamma^2 + bn\sigma_{\alpha\gamma}^2 + an\sigma_{\beta\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \sigma_{(\gamma)}^2 = 0$
$E(MS_{AB}) = \frac{nc}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{(ij)}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : (\alpha\beta)_{(ij)} = 0$
$E(MA_{AC}) = bn\sigma_{\alpha\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \sigma_{\alpha\gamma}^2 = 0$
$E(MS_{BC}) = an\sigma_{\beta\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \sigma_{\beta\gamma}^2 = 0$
$E(MS_{ABC}) = n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$	$H_0 : \sigma_{\alpha\beta\gamma}^2 = 0$
$E(MS_E) = \sigma_\epsilon^2$	

error term by the addition or substraction of several mean squares whose expected value is $bn\sigma_{\alpha\gamma}^2 + an\sigma_{\beta\gamma}^2 + n\sigma_{\alpha\beta\gamma}^2 + \sigma_\epsilon^2$, which is equal to the expected mean square of C under null hypothesis $H_0 : \sigma_\gamma^2 = 0$. The resulting ratio, $MS_C/[MS_{AC} + MS_{BC} - MS_{ABC}]$, has, under the null hypothesis, an approximate F-distribution with $c - 1$ and ν degrees of freedom, where ν is given by

$$\nu = \frac{(MS_{AC} + MS_{BC} - MS_{ABC})^2}{\frac{(MS_{AC})^2}{(a-1)(c-1)} + \frac{(MS_{BC})^2}{(b-1)(c-1)} + \frac{(-MS_{ABC})^2}{(a-1)(b-1)(c-1)}}. \quad (13.39)$$

If the computed value of ν is not an integer, round down to the nearest integer. Such an approximation is done according to the method of Satterthwaite's approximation. For more details concerning this approximation, see, for example, Khuri (2010, Chapter 9) and Dean and Voss (1999, p. 117).

Example 13.8 *Consider the following balanced model,*

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{(j)} + (\alpha\beta)_{(ij)} + \delta_{j(k)} + (\alpha\delta)_{j(ik)} + \epsilon_{ijk(l)},$$
$$i = 1,2,3;\ j = 1,2,3,4;\ k = 1,2,3;\ l = 1,2. \quad (13.40)$$

TABLE 13.7 Test Statistics for Model (13.38)

H_0	F-Test Statistic	α-Critical Value
$\alpha_{(i)} = 0$	MS_A/MS_{AC}	$F_{\alpha,a-1,(a-1)(c-1)}$
$\beta_{(j)} = 0$	MS_B/MS_{BC}	$F_{\alpha,b-1,(b-1)(c-1)}$
$\sigma_\gamma^2 = 0$	$MS_C/[MS_{AC} + MS_{BC} - MS_{ABC}]$	$F_{\alpha,c-1,\nu}$
$(\alpha\beta)_{(ij)} = 0$	MS_{AB}/MS_{ABC}	$F_{\alpha,(a-1)(b-1),(a-1)(b-1)(c-1)}$
$\sigma_{\alpha\gamma}^2 = 0$	MS_{AC}/MS_{ABC}	$F_{\alpha,(a-1)(c-1),(a-1)(b-1)(c-1)}$
$\sigma_{\beta\gamma}^2 = 0$	MS_{BC}/MS_{ABC}	$F_{\alpha,(b-1)(c-1),(a-1)(b-1)(c-1)}$
$\sigma_{\alpha\beta\gamma}^2 = 0$	MS_{ABC}/MS_E	$F_{\alpha,(a-1)(b-1)(c-1),abc(n-1)}$

TABLE 13.8 ANOVA Table for Example 13.8

Source	DF	MS	E(MS)
A	2	34.65	$24\,\sigma_{\alpha}^2 + 6\sigma_{\alpha\beta}^2 + 2\sigma_{\alpha\delta(\beta)}^2 + \sigma_{\epsilon}^2$
B	3	615.98	$18\,\sigma_{\beta}^2 + 6\sigma_{\alpha\beta}^2 + 6\sigma_{\delta(\beta)}^2 + 2\sigma_{\alpha\delta(\beta)}^2 + \sigma_{\epsilon}^2$
$A * B$	6	11.49	$6\,\sigma_{\alpha\beta}^2 + 2\sigma_{\alpha\delta(\beta)}^2 + \sigma_{\epsilon}^2$
$C(B)$	8	17.92	$6\,\sigma_{\delta(\beta)}^2 + 2\sigma_{\alpha\delta(\beta)}^2 + \sigma_{\epsilon}^2$
$A * C(B)$	16	6.25	$2\,\sigma_{\alpha\delta(\beta)}^2 + \sigma_{\epsilon}^2$
Error	36	2.98	σ_{ϵ}^2

TABLE 13.9 Test Statistics for Model (13.40)

H_0	F-Test Statistic	5%-Critical Value
$\sigma_{\alpha}^2 = 0$	$MS_A/MS_{AB} = 3.016$	$F_{0.05,2,6} = 5.14$
$\sigma_{\beta}^2 = 0$	$MS_B/[MS_{AB} + MS_{C(B)} - MS_{A*C(B)}] = 26.597$	$F_{0.05,3,8} = 4.07$
$\sigma_{\alpha\beta}^2 = 0$	$MS_{AB}/MS_{A*C(B)} = 1.838$	$F_{0.05,6,16} = 2.74$
$\sigma_{\delta(\beta)}^2 = 0$	$MS_{C(B)}/MS_{A*C(B)} = 2.867$	$F_{0.05,8,16} = 2.59$
$\sigma_{\alpha\delta(\beta)}^2 = 0$	$MS_{A*C(B)}/MS_E = 2.097$	$F_{0.05,16,36} = 1.92$

This model concerns an experiment involving three factors, A, B, C with A and B crossed and C is nested within B. Two replicated observations are available for each combination of the levels of A, B, and C. Hence, the population structure is $[(i)(j : k)] : l$. *All the effects are random and assumed to be independently distributed as normal variates with zero means and variances,* $\sigma_{\alpha}^2, \sigma_{\beta}^2, \sigma_{\alpha\beta}^2, \sigma_{\delta(\beta)}^2, \sigma_{\alpha\delta(\beta)}^2, \sigma_{\epsilon}^2$. *Based on a data set, the corresponding ANOVA table along with the expected mean squares are shown in Table 13.8.*

On the basis of the expected mean squares expressions, tests concerning the various effects are given in Table 13.9. Note that the test concerning $H_0 : \sigma_{\beta}^2 = 0$ *requires using Satterthwaite's approximation since the denominator of the approximate F-statistic is a linear combination of mean squares, namely,* $MS_{AB} + MS_{C(B)} - MS_{A*C(B)}$. *The number of degrees of freedom for the denominator is*

$$\nu = \frac{[MS_{AB} + MS_{C(B)} - MS_{A*C(B)}]^2}{\frac{(MS_{AB})^2}{6} + \frac{(MS_{C(B)})^2}{8} + \frac{(-MS_{A*C(B)})^2}{16}}$$

This gives $\nu = 8.305$, *which is rounded down to the nearest integer, namely, 8.*

REFERENCES

Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*. Springer-Verlag, New York.

Khuri, A. I. (1982). Direct products: A powerful tool for the analysis of balanced data. *Comm. Stat. Theory Methods*, 11, 2903–2920.

Khuri, A. I. (2010). *Linear Model Methodology*. Chapman & Hall/CRC, Boca Raton, FL.

Zyskind, G. (1962). On structure, relation, **Σ**, and expectation of mean squares. *Sankhyā Ser. A*, 24, 115–148.

EXERCISES

13.1 Let A and B be two factors with levels $a_1, a_2, \dots, a_p$ for A and levels $b_1, b_2, \dots, b_q$ for B. Factor A represents different drugs and B represents time. A random sample of size n from a common population of subjects was chosen. Each subject was given level a_1 of A and then observed under the q levels of B. Another random sample of size n of subjects was chosen and the subjects were each given level a_2 of A and then observed under the q levels of B. The experiment continued by selecting other groups of subjects, each of size n, with each group receiving one level A and then observed under the q levels of B until all levels of A have been used. Let y_{ijk} denote a measurement on subject k in the ith group $(i = 1, 2, \dots, p)$ under the treatment combination $a_i b_j$. The subject effect is random and is assumed to be normally distributed with a zero mean and variance $\sigma^2_{\pi(\alpha)}$, independently of the error term, ϵ_{ijk}, which has the normal distribution $N(0, \sigma^2_\epsilon)$.

(a) Write down the population structure for this experiment.

(b) Write down the complete model.

(c) Give the corresponding ANOVA table including the expected mean squares expressions.

13.2 An experiment was conducted to test the effect of a certain treatment on skin sensitivity. Three subjects were given a placebo (control group) and another group of three subjects were given the treatment. Each subject in a group was administered the products on two different locations of the body (location 1 and location 2). Duplicate tests were run. The resulting data set is given in the table below.

		Location (B)	
Group (A)	Subject	Location 1	Location 2
Control	1	70, 72	88, 93
	2	65, 64	80, 82
	3	70, 69	75, 73
Treatment	4	75, 73	88, 87
	5	66, 71	68, 75
	6	75, 76	95, 94

(a) Give the population structure.

(b) Write down the complete model.

(c) Give the ANOVA table including the expected mean squares expressions.

(d) Provide test statistics for the various effects in the model and specify the corresponding p-values.

Note: The SAS statements needed for this exercise are given below.

```
DATA;
INPUT A C B y@@;
CARDS;
```

```
(enter here the data according to the INPUT statement)
PROC PRINT;
PROC GLM;
CLASS A B C;
MODEL y=A C(A) B A*B B*C(A);
RANDOM C(A) B*C(A)/TEST;
RUN;
```

Factor C refers to subjects. An explanation of these SAS statements will be given in Chapter 15 of Part 3 of this book.

13.3 *Satterthwaite's approximation* states that if $MS_1, MS_2, \ldots, MS_k$ are mutually independent mean squares such that $m_i MS_i/\delta_i \sim \chi^2_{m_i}$, where $\delta_i = E(MS_i)$, $i = 1, 2, \ldots, k$, and if $MS^* = \sum_{i=1}^k a_i MS_i$, where $a_1, a_2, \ldots, a_k$ are known nonzero constants, then $\nu MS^*/\delta$ has approximately the chi-squared distribution with ν degrees of freedom, where $\delta = E(MS^*)$ and ν is given by the formula

$$\nu = \frac{(\sum_{i=1}^k a_i MS_i)^2}{\sum_{i=1}^k a_i^2 MS_i^2/m_i}.$$

Use this approximation and the data set in Exercise 2 to find an approximate 95% confidence interval on the variance component $\sigma^2_{\beta\gamma(\alpha)}$, which is associated with the interaction effect, $B * C(A)$, of location and subjects within treatments.

13.4 Consider again the data set in Exercise 2. Show how to obtain a 95% confidence interval on the ratio

$$\rho = \frac{\sigma^2_{\beta\gamma(\alpha)}}{\sigma^2_\epsilon + \sigma^2_{\beta\gamma(\alpha)}}.$$

13.5 Consider the balanced three-fold nested model

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{i(j)} + \gamma_{ij(k)} + \epsilon_{ijk(l)},$$

where $\alpha_{(i)}$ is fixed and $\beta_{i(j)}$ and $\gamma_{ij(k)}$ are random ($i = 1, 2, ..., a$; $j = 1, 2, ..., b$; $k = 1, 2, ..., c$; $l = 1, 2, ..., n$). It is assumed that $\beta_{i(j)}$, $\gamma_{ij(k)}$, and $\epsilon_{ijk(l)}$ are independently distributed as normal variables with zero means and variances $\sigma^2_{\beta(\alpha)}$, $\sigma^2_{\gamma(\alpha\beta)}$, and σ^2_ϵ, respectively.

(a) Describe the population structure.

(b) Give the expected mean squares for all the effects in the ANOVA table.

(c) Give an expression for the probability of making a Type II error concerning the hypothesis $H_0 : \alpha_{(i)} = 0$ for all i, for a 5% level of significance.

(d) Let $\hat{\sigma}^2_{\beta(\alpha)}$ denote the ANOVA estimator of $\sigma^2_{\beta(\alpha)}$. Give an expression for computing the probability that $\hat{\sigma}^2_{\beta(\alpha)} < 0$.

13.6 Consider the following ANOVA table for a balanced random three-way classification model (the usual assumptions of normality and independence of the random effects are assumed valid):

Source	DF	MS	E(MS)
A	4	37.25	$12\sigma_{\alpha}^2 + 3\sigma_{\alpha\beta}^2 + 4\sigma_{\alpha\gamma}^2 + \sigma_{\epsilon}^2$
B	3	66.82	$15\sigma_{\beta}^2 + 3\sigma_{\alpha\beta}^2 + 5\sigma_{\beta\gamma}^2 + \sigma_{\epsilon}^2$
A*B	12	10.65	$3\sigma_{\alpha\beta}^2 + \sigma_{\epsilon}^2$
C	2	89.16	$20\sigma_{\gamma}^2 + 4\sigma_{\alpha\gamma}^2 + 5\sigma_{\beta\gamma}^2 + \sigma_{\epsilon}^2$
A*C	8	5.19	$4\sigma_{\alpha\gamma}^2 + \sigma_{\epsilon}^2$
B*C	6	12.05	$5\sigma_{\beta\gamma}^2 + \sigma_{\epsilon}^2$
Error	24	3.08	σ_{ϵ}^2

Test the hypothesis

$$H_0 : E(MS_{\alpha}) + E(MS_E) = E(MS_{\alpha\beta}) + E(MS_{\alpha\gamma}) \quad \text{versus}$$
$$H_a : E(MS_{\alpha}) + E(MS_E) > E(MS_{\alpha\beta}) + E(MS_{\alpha\gamma}),$$

at the 10% level of significance.

13.7 Consider the balanced random one-way classification model,

$$y_{ij} = \mu + \alpha_{(i)} + \epsilon_{i(j)}, \quad i = 1, 2, \ldots, a; j = 1, 2, \ldots, n.$$

The usual assumptions concerning the random effects are assumed valid with variance components, σ_{α}^2 and σ_{ϵ}^2.

(a) Obtain a $(1 - \alpha)100\%$ confidence interval on $\sigma_{\alpha}^2/\sigma_{\epsilon}^2$.

(b) Obtain a $(1 - \alpha)100\%$ confidence interval on $\sigma_{\epsilon}^2/(\sigma_{\alpha}^2 + \sigma_{\epsilon}^2)$.

13.8 Consider the model,

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{i(j)} + \gamma_{i(k)} + \delta_{ik(l)} + (\beta\gamma)_{i(jk)} + \epsilon_{ik(jl)},$$

$i = 1, 2, 3$; $j = 1, 2, 3$; $k = 1, 2, 3, 4$; $l = 1, 2, 3$, where $\alpha_{(i)}$ is fixed; $\beta_{i(j)}$, $\gamma_{i(k)}$, and $\delta_{ik(l)}$ are random. It is assumed that the random effects are independently distributed as normal variates with zero means and variances, $\sigma_{\beta(\alpha)}^2$, $\sigma_{\gamma(\alpha)}^2$, $\sigma_{\delta(\alpha\gamma)}^2$, $\sigma_{\beta\gamma(\alpha)}^2$, and σ_{ϵ}^2.

(a) Give the population structure.

(b) Write down the ANOVA table including the expected mean squares.

(c) Give the test statistic for testing the hypothesis $H_0 : \sigma_{\delta(\alpha\gamma)}^2 = 2\sigma_{\epsilon}^2$ versus $H_a : \sigma_{\delta(\alpha\gamma)}^2 \neq 2\sigma_{\epsilon}^2$.

13.9 Consider again Exercise 8.

(a) Obtain a $(1-\alpha)100\%$ confidence region on the parameter vector $\theta = (\theta_1, \theta_2)'$, where θ_1 and θ_2 are the expected values of the mean squares for $B(A)$ and $B * C(A)$, respectively.

(b) Let $\hat{\sigma}^2_{\beta(\alpha)}$ be the ANOVA estimator of $\sigma^2_{\beta(\alpha)}$. Give an expression that can be used to compute the probability $P(\hat{\sigma}^2_{\beta(\alpha)} < 0)$ in terms of $\Delta = \sigma^2_{\beta(\alpha)}/[3\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_{\epsilon}]$.

13.10 Consider a balanced data situation involving three factors, A, B, and C. Factors A and B are crossed, and C is nested within A and B (that is, within the AB subclasses). Two replicated observations were obtained for each combination of the levels of A, B, and C. Factor A has three levels and factors B and C have 4 levels each. The three levels of A are of particular interest whereas the levels of B and C were selected at random. All random effects are assumed to be independent and have the normal distribution with zero means and corresponding variance components.

(a) Give the corresponding population structure. Use subscripts i, j, k for factors A, B, and C, respectively, and subscript l for the replications.

(b) List all the admissible means and corresponding components, then write down the model for this experiment.

(c) Let $\sigma^2_{\alpha\beta}$ denote the variance component for the $A * B$ interaction. Let $\hat{\sigma}^2_{\alpha\beta}$ be its ANOVA estimator. Give an expression for $\text{var}(\hat{\sigma}^2_{\alpha\beta})$.

(d) Let $\alpha_{(i)}$ denote the effect of level i of factor A $(i = 1, 2, 3)$. Obtain a $(1-\alpha)100\%$ confidence interval on $\alpha_{(1)} - \alpha_{(2)}$.

13.11 Consider model (13.22). Let $\boldsymbol{P}_i$ be the matrix corresponding to the ith sum of squares in the ANOVA table for this model $(i = 1, 2, \ldots, q+1)$ whose rank is m_i.

(a) Show that $\boldsymbol{X} = \sum_{i=0}^{q-p} \boldsymbol{P}_i \boldsymbol{X}$.

(b) Show that no linear relationships exist among $\boldsymbol{P}_0, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_{q+1}$.

(c) Show that the rank of $\sum_{i=0}^{q-p} \boldsymbol{P}_i$ is equal to $\sum_{i=0}^{q-p} m_i$.

14

Multiresponse Models

The experimental situations encountered so far in this book concerned experiments in which a single response variable was observed. Quite often, however, several response variables may be of interest in a given experiment. Such response variables may represent several attributes of a certain product, or several desirable characteristics the product is required to have. By definition, an experiment in which a number of response variables are measured for each setting of a group of control variables is called a *multiresponse experiment*. It is quite common in experimental work to observe several response variables rather than a single response. For example, in the food industry, a product may be evaluated on the basis of acceptability, nutritional value, economics, and other considerations. In this case, the desirability of the product depends on finding the combinations of the various ingredients of the product that enhance the quality of such attributes.

The multivariate nature of a multiresponse experiment should be considered in any analysis of data from such an experiment. It is obvious that univariate analyses applied to the individual responses should not be considered since they do not take into account any interrelationships that may exist among the responses. For example, optimum conditions on the control variables that maximize (or minimize) a single response variable in a multiresponse situation may be far from optimal for the other responses. Optimum conditions should therefore be determined in a manner that make them favorable to all the responses in the multiresponse system.

In this chapter we provide coverage of some multiresponse techniques. These include the estimation of model parameters and the testing of lack of fit of a multiresponse model.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

14.1 MULTIRESPONSE ESTIMATION OF PARAMETERS

Suppose that we have r response variables that can be measured for each setting of k control variables denoted by $x_1, x_2, \ldots, x_k$. Let n be the number of experimental runs that are available on the r responses. At the uth run, the settings of the control variables are denoted by x_{ui} ($i = 1, 2, \ldots, k$; $u = 1, 2, \ldots, n$). Let $y_{u1}, y_{u2}, \ldots, y_{ur}$ represent the corresponding values of the responses. Suppose that these responses are related to control variables through the models,

$$y_{ui} = f_i(\boldsymbol{x}_u, \boldsymbol{\beta}) + \epsilon_{ui}, \quad i = 1, 2, \ldots r; u = 1, 2, \ldots, n, \tag{14.1}$$

where $\boldsymbol{x}_u = (x_{u1}, x_{u2}, \ldots, x_{uk})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a vector of p unknown parameters, ϵ_{ui} is a random experimental error, and f_i is a known function. The r functions in (14.1) can be represented using a single matrix equation of the form

$$\boldsymbol{Y} = \mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{14.2}$$

where $\boldsymbol{Y}$ is a matrix of order $n \times r$ whose ith column is $\boldsymbol{y}_i$, the ith response vector of observations ($i = 1, 2, \ldots, r$), $\boldsymbol{\epsilon}$ is a matrix of order $n \times r$ whose ith column is $\boldsymbol{\epsilon}_i$, the random error vector for the ith response ($i = 1, 2, \ldots, r$), $\boldsymbol{D}$ denotes the design matrix which is of order $n \times k$ and its n rows are $\boldsymbol{x}'_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_n$, and $\mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta})$ is a matrix of order $n \times r$ whose ith column consists of the values of $f_i(\boldsymbol{x}_u, \ \boldsymbol{\beta})$ for $u = 1, 2, \ldots, n$ ($i = 1, 2, \ldots, r$). The rows of $\boldsymbol{\epsilon}$ are assumed to be independently and identically distributed as $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an unknown variance-covariance matrix for the r responses.

Box and Draper (1965) used a Bayesian approach for estimating $\boldsymbol{\beta}$. By assuming a noninformative prior distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, they showed that an estimate of $\boldsymbol{\beta}$ can be obtained by finding the value of $\boldsymbol{\beta}$ that minimizes the determinant

$$g(\boldsymbol{\beta}) = |[\boldsymbol{Y} - \mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta})]'[\boldsymbol{Y} - \mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta})]|. \tag{14.3}$$

This estimation method is called the *Box–Draper determinant criterion.*

It should be noted that the Box–Draper criterion can lead to meaningless results when linear relationships exist among the responses. To explain this, suppose that we have the following m linear relationships among the responses:

$$\boldsymbol{B}\boldsymbol{Y}_u = \boldsymbol{c}, \quad u = 1, 2, \ldots, n, \tag{14.4}$$

where $\boldsymbol{Y}'_u$ is the uth row of $\boldsymbol{Y}$, $\boldsymbol{B}$ is an $m \times r$ matrix of rank m ($< r$) of constant coefficients, and $\boldsymbol{c}$ is a vector of m constants. The n equalities in (14.4) can be represented in matrix form as

$$\boldsymbol{B}\boldsymbol{Y}' = \boldsymbol{1}'_n \otimes \boldsymbol{c}. \tag{14.5}$$

Taking the expected values of both sides of (14.5) we get

$$\boldsymbol{B}\mathcal{F}'(\boldsymbol{D}, \boldsymbol{\beta}) = \boldsymbol{1}'_n \otimes \boldsymbol{c}. \tag{14.6}$$

From (14.5) and (14.6) we conclude

$$\boldsymbol{B}[\boldsymbol{Y}' - \mathcal{F}'(\boldsymbol{D}, \boldsymbol{\beta})] = \boldsymbol{0}. \tag{14.7}$$

This indicates the presence of m linearly independent relationships among the rows of the matrix $\boldsymbol{Y}' - \mathcal{F}'(\boldsymbol{D}, \boldsymbol{\beta})$. Hence, the matrix $[\boldsymbol{Y} - \mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta})]'[\boldsymbol{Y} - \mathcal{F}(\boldsymbol{D}, \boldsymbol{\beta})]$ must be singular and its determinant in (14.3) must be equal to zero for all values of $\boldsymbol{\beta}$. Thus it will be meaningless to minimize the determinant in this case.

Roundoff error can cause the determinant in (14.3) to be different from zero even in the presence of linear relationships among the responses. Box et al. (1973) proposed that the multiresponse data should be checked first for the presence of linear relationships before applying the Box–Draper determinant criterion. This is accomplished by computing the eigenvalues of the matrix $\boldsymbol{G}\boldsymbol{G}'$, where $\boldsymbol{G}$ is an $r \times n$ matrix of the form

$$\boldsymbol{G} = \boldsymbol{Y}'\left[\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right], \tag{14.8}$$

where $\boldsymbol{J}_n$ is a matrix of ones of order $n \times n$. It can be shown that m linearly independent relationships exist among the responses if and only if $\boldsymbol{G}\boldsymbol{G}'$ has a zero eigenvalue of multiplicity m [see Khuri (1996, Section 3.2); Khuri and Cornell (1996, Section 7.2.2)]. Roundoff error can also cause the eigenvalues of $\boldsymbol{G}\boldsymbol{G}'$ to be different from zero even when linear relationships exist among the responses. This situation can be handled by having small eigenvalues of $\boldsymbol{G}\boldsymbol{G}'$ subjected to further analysis to determine if they correspond to a zero eigenvalue. A discussion concerning the implementation of such analysis is described in Khuri and Cornell (1996, pp. 256–257). See also Khuri (1996, pp. 382–383). The following is a brief account of this analysis:

Suppose that λ is a "small" eigenvalue of $\boldsymbol{G}\boldsymbol{G}'$ which would be considered equal to zero if it were not for the roundoff errors. If such errors are distributed independently and uniformly over the interval $(-\delta, \delta)$, where δ is equal to one-half of the last digit recorded if all the response values are rounded off to the same number of significant figures, then the expected value of λ is approximately equal to

$$\mathrm{E}(\lambda) = (n-1)\sigma_{re}^2, \tag{14.9}$$

where $\sigma_{re}^2 = \delta^2/3$ is the roundoff variance. Furthermore, the variance of λ has an upper bound that is approximately given by

$$\mathrm{var}(\lambda) \leq \left[\frac{9nr}{5} + nr(nr-1) - (n-1)^2\right]\sigma_{re}^4. \tag{14.10}$$

The inequality in (14.10) was developed by Khuri and Conlon (1981). Using (14.9) and (14.10) it can be determined that if a small eigenvalue of $\boldsymbol{G}\boldsymbol{G}'$ falls within two or three standard deviation from $\mathrm{E}(\lambda)$, then it can be considered as equal to zero. A conservative value for the standard deviation is taken to be equal to the square root of the upper bound in (14.10).

If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{G}\boldsymbol{G}'$ corresponding to an eigenvalue considered equal to zero, then we have

$$\boldsymbol{G}\boldsymbol{G}'\boldsymbol{v} = \boldsymbol{0},$$

which is equivalent to

$$\boldsymbol{G}'\boldsymbol{v} = \boldsymbol{0}.$$

Using (14.8), we then have

$$\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{Y}\boldsymbol{v} = \boldsymbol{0},$$

which can be expressed as

$$\boldsymbol{Y}\boldsymbol{v} = \gamma\boldsymbol{1}_n,$$

where $\gamma = \frac{1}{n}\boldsymbol{1}_n'\boldsymbol{Y}\boldsymbol{v}$. We thus see that the elements of $\boldsymbol{v}$ define a linear relationship among the r responses of the form

$$\sum_{i=1}^{r} Y_{ui}v_i = \gamma, \quad u = 1, 2, \ldots, n,$$

where Y_{ui} is the (u, i)th element of $\boldsymbol{Y}$ and v_i is the ith element of $\boldsymbol{v}$. The constant γ is the same in all n observations on the responses.

For example, suppose that in a multiresponse situation we have $r = 4$ responses with $n = 10$ observations from each response. The multiresponse data are rounded off to one decimal place. Then, $\delta = 0.05$, $\sigma_{re}^2 = \delta^2/3 = 0.0008$, and from (14.9), $\mathrm{E}(\lambda) = 0.0072$. The square root of the upper bound in (14.10) is equal to 0.0315, which represents a conservative estimate of the standard deviation of λ. The eigenvalues of $\boldsymbol{G}\boldsymbol{G}'$ are $\lambda_1 = 0.0172$, $\lambda_2 = 0.98$, $\lambda_3 = 21.9$, $\lambda_4 = 98.65$. We note that λ_1 falls within one standard deviation from $\mathrm{E}(\lambda)$ whereas the next smallest eigenvalue, namely λ_2, is more that 30 standard deviation away. We conclude that λ_1 corresponds to a zero eigenvalue of $\boldsymbol{G}\boldsymbol{G}'$ of multiplicity one.

14.2 LINEAR MULTIRESPONSE MODELS

This is a special case of multiresponse models. By definition, a linear multiresponse model is a multiresponse model that is linear in the parameters. The model for the ith response is of the form

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, r, \tag{14.11}$$

where $\boldsymbol{y}_i$ is a vector of observations from the ith response, $\boldsymbol{X}_i$ is a matrix of order $n \times p_i$ and rank p_i, and $\boldsymbol{\beta}_i$ is a vector of unknown parameters. The models in (14.11) can be represented as a single multiresponse model of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{14.12}$$

where

$$\boldsymbol{y} = (\boldsymbol{y}_1' : \boldsymbol{y}_2' : \ldots : \boldsymbol{y}_r')', \boldsymbol{\varepsilon} = \left(\boldsymbol{\epsilon}_1' : \boldsymbol{\epsilon}_2' : \ldots : \boldsymbol{\epsilon}_r'\right)',$$

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1' : \boldsymbol{\beta}_2' : \ldots : \boldsymbol{\beta}_r'\right)',$$

and $\boldsymbol{X}$ is a block-diagonal matrix of the form $\boldsymbol{X} = \text{diag}(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_r)$. This model is called a *multiple design multivariate linear model* (see McDonald, 1975).

If, as before, the assumptions concerning the distribution of the rows of the error matrix $\boldsymbol{\epsilon}$ in (14.2) are valid, then $\boldsymbol{\varepsilon}$ has the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix for the r responses. It follows that *best linear unbiased estimator* of $\boldsymbol{\beta}$ is the generalized least-squares estimator given by

$$\hat{\boldsymbol{\beta}} = \left[\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\right]^{-1} \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}. \tag{14.13}$$

Since $\boldsymbol{\Sigma}$ is generally unknown, an estimator of $\boldsymbol{\Sigma}$ must be used. One such estimator was proposed by Zellner (1962) and is given by $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij})$, where

$$\hat{\sigma}_{ij} = \frac{1}{n}\boldsymbol{y}_i'\left[\boldsymbol{I}_n - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\right]\left[\boldsymbol{I}_n - \boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j'\right]\boldsymbol{y}_j, \quad i,j = 1, 2, \ldots, r.$$

Replacing $\boldsymbol{\Sigma}$ with this estimator in (14.13) results in a so-called estimated generalized least-squares estimate of $\boldsymbol{\beta}$ and is denoted by $\hat{\boldsymbol{\beta}}_e$. This estimator, however, does have the optimal BLUE property that $\hat{\boldsymbol{\beta}}$ has. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is approximately given by

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_e) \approx \left[\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\right]^{-1}.$$

The ith predicted response value at a point $\boldsymbol{x}$ in a region of interest is

$$\hat{y}_i(\boldsymbol{x}) = \boldsymbol{f}_i'(\boldsymbol{x})\hat{\boldsymbol{\beta}}_{ie}, \quad i = 1, 2, \ldots, r, \tag{14.14}$$

where $\boldsymbol{f}_i'(\boldsymbol{x})$ has the same form as a row of $\boldsymbol{X}_i$ but is evaluated at $\boldsymbol{x}$, and $\hat{\boldsymbol{\beta}}_{ie}$ is the ith portion of $\hat{\boldsymbol{\beta}}_e = (\hat{\boldsymbol{\beta}}_{1e}', \hat{\boldsymbol{\beta}}_{2e}', \ldots, \hat{\boldsymbol{\beta}}_{re}')'$. The polynomial in (14.14) is of degree $d_i(\geq 1)$, $i = 1, 2, \ldots, r$. The models in (14.14) can then be expressed as a single matrix model of the form

$$\hat{\boldsymbol{y}}(\boldsymbol{x}) = \left[\oplus_{i=1}^{m} \boldsymbol{f}_i'(\boldsymbol{x})\right]\hat{\boldsymbol{\beta}}_e, \tag{14.15}$$

where $\hat{\boldsymbol{y}}(\boldsymbol{x})$ is a column vector whose ith element is $\hat{y}_i(\boldsymbol{x})$, $i = 1, 2, \ldots, r$. The estimated variance-covariance matrix of $\hat{\boldsymbol{y}}(\boldsymbol{x})$ is approximately given by

$$\widehat{var}\left[\hat{\boldsymbol{y}}(\boldsymbol{x})\right] \approx [\oplus_{i=1}^{m}\boldsymbol{f}_i'(\boldsymbol{x})][\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}]^{-1}[\oplus_{i=1}^{m}\boldsymbol{f}_i(\boldsymbol{x})]. \tag{14.16}$$

In the special case when the $\boldsymbol{X}_i's$ are all equal to $\boldsymbol{X}_0$, then it can be shown (see Exercise 14.5) that $\hat{\boldsymbol{\beta}}$ takes the simpler form

$$\hat{\boldsymbol{\beta}} = \left[\boldsymbol{I}_r \otimes \left(\boldsymbol{X}_0'\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0'\right]\boldsymbol{y}. \tag{14.17}$$

This estimator does not depend $\boldsymbol{\Sigma}$ and is the same as the ordinary least-squares estimator of $\boldsymbol{\beta}$ which results from fitting the r response models individually.

14.3 LACK OF FIT OF A LINEAR MULTIRESPONSE MODEL

A linear multiresponse model is said to suffer from *lack of fit* if at least one of its univariate models does not adequately represent the true mean of its corresponding response vector over the experimental region. Since the responses can be correlated, lack of fit in at least one univariate model can influence the fit of the remaining models. It is therefor necessary to recognize the multivariate nature of the multiresponse data in the development of a lack of fit test for a linear multiresponse model.

The r univariate models in (14.11) can be combined using a single model of the form

$$\boldsymbol{Y} = \mathcal{X}\mathcal{B} + \boldsymbol{\epsilon}, \tag{14.18}$$

where $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are the same as in model (14.2), $\mathcal{X} = [\boldsymbol{X}_1 : \boldsymbol{X}_2 : \ldots : \boldsymbol{X}_r]$, and $\mathcal{B} = \oplus_{i=1}^{r} \boldsymbol{\beta}_i$ is a block-diagonal matrix. The matrix $\mathcal{X}$ is not necessarily of full column rank since some of the control variables and possibly their powers or mixed products can appear in more than one model. If ρ denotes the rank of $\mathcal{X}$, then $\rho \leq \sum_{i=1}^{r} p_i$, where, if we recall, p_i is the number of columns of $\boldsymbol{X}_i$, $i = 1, 2, \ldots, r$. It is assumed that the rows of $\boldsymbol{\epsilon}$ are independent vectors from multivariate normal distributions with zero means and a common nonsingular variance-covariance matrix $\boldsymbol{\Sigma}$ of order $r \times r$. It is also assumed that there are replicated runs on all the responses at some points in the experimental region. The replicated runs are considered to be taken at each of the first n_0 points of the design ($1 \leq n_0 < n$).

14.3.1 The Multivariate Lack of Fit Test

The development of the multivariate lack of fit test described here is based on a method proposed by Khuri (1985). This method amounts to introducing a single response which consists of an arbitrary linear combination of the responses whose corresponding model can be tested for lack of fit using the familiar univariate lack of fit test. For this purpose, let $\boldsymbol{c} = (c_1, c_2, \ldots, c_r)'$ be an arbitrary nonzero vector of r elements. Multiplying both sides of the model in (14.18) on the right by $\boldsymbol{c}$, we get

$$\boldsymbol{y_c} = \mathcal{X}\boldsymbol{\beta_c} + \boldsymbol{\epsilon}_c, \tag{14.19}$$

where $\boldsymbol{y_c} = \boldsymbol{Y}\boldsymbol{c}$, $\boldsymbol{\beta_c} = \mathcal{B}\boldsymbol{c}$, $\boldsymbol{\epsilon}_c = \boldsymbol{\epsilon}\,\boldsymbol{c}$. Since $\boldsymbol{y_c}$ is a linear combination of normally distributed vectors, it must have the multivariate normal distribution with a variance-covariance matrix given by

$$\text{var}(\boldsymbol{y_c}) = \sigma_{\boldsymbol{c}}^2 \boldsymbol{I}_n, \tag{14.20}$$

where $\sigma_{\boldsymbol{c}}^2 = \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}$.

Using *Roy's union-intersection principle* of multivariate analysis [see, e.g., Roy (1953), Morrison (1976), Chapters 4 and 5], it can be stated that the linear multiresponse model in (14.18) is correct if and only if the univariate models in in (14.19) are correct for all $\boldsymbol{c} \neq \boldsymbol{0}$. This is equivalent to stating that model (14.18) is considered incorrect if and only for at least one $\boldsymbol{c} \neq \boldsymbol{0}$ the model in (14.19) is incorrect. Since the model for $\boldsymbol{y_c}$ is univariate with a variance of the form given in (14.20), we can test its adequacy by using the standard procedure for testing its lack of fit, given the availability of replicated observations on $\boldsymbol{y_c}$ at each of the first n_0 design points. This standard procedure is described, for example, in

[Draper and Smith (1981), Section 1.5, and Khuri and Cornell (1996), Section 2.6]. It is based on partitioning the residual sum of squares, denoted by $SS_E(\boldsymbol{c})$, which results from fitting the univariate model in (14.19), into two sums of squares, namely, the lack of fit sum of squares, denoted by $SS_{LOF}(\boldsymbol{c})$, and the pure error sum of squares, denoted by $SS_{PE}(\boldsymbol{c})$. The latter sum of squares is computed by using the replicated observations on the univariate response $\boldsymbol{y_c}$. The aforementioned residual sum of squares is given by

$$SS_E(\boldsymbol{c}) = \boldsymbol{y}'_{\boldsymbol{c}}[\boldsymbol{I}_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-}\mathcal{X}']\boldsymbol{y_c}. \tag{14.21}$$

Note that a generalized inverse of $\mathcal{X}'\mathcal{X}$ is used (14.21) since, as was mentioned earlier, $\mathcal{X}$ is not necessarily of full column rank. It may also be recalled from Theorem 8.2(iii) that $\mathcal{X}(\mathcal{X}'\mathcal{X})^{-}\mathcal{X}'$ is invariant to the choice of this generalized inverse. Using the replicated observations on the univariate response $\boldsymbol{y_c}$ at the first n_0 design points, the pure error sum of squares is computed using the formula

$$SS_{PE}(\boldsymbol{c}) = \boldsymbol{y}'_{\boldsymbol{c}}\boldsymbol{K}\boldsymbol{y_c}, \tag{14.22}$$

where

$$\boldsymbol{K} = \text{diag}(\boldsymbol{K}_1, \boldsymbol{K}_2, \dots, \boldsymbol{K}_{n_0}, \boldsymbol{0}) \tag{14.23}$$

is order $n \times n$. Let v_i be the number of repeated runs at the ith design point ($i = 1, 2, \dots, n_0$). The zero matrix in (14.23) is of order $(n - \sum_{i=1}^{n_0} v_i) \times (n - \sum_{i=1}^{n_0} v_i)$ and $\boldsymbol{K}_i$ is a matrix of order $v_i \times v_i$ of the form

$$\boldsymbol{K}_i = \boldsymbol{I}_{v_i} - \frac{1}{v_i}\boldsymbol{J}_{v_i}, \quad i = 1, 2, \dots, n_0, \tag{14.24}$$

where $\boldsymbol{J}_{v_i}$ is the matrix of ones of order $v_i \times v_i$. The number of degrees of freedom for $SS_{PE}(\boldsymbol{c})$ is denoted by v_{PE} and is equal to

$$v_{PE} = \sum_{i=1}^{n_0}(v_i - 1). \tag{14.25}$$

The lack of fit sum of squares, $SS_{LOF}(\boldsymbol{c})$, is computed from subtracting $SS_{PE}(\boldsymbol{c})$ from $SS_E(\boldsymbol{c})$, that is,

$$SS_{LOF}(\boldsymbol{c}) = \boldsymbol{y}'_{\boldsymbol{c}}[\boldsymbol{I}_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-}\mathcal{X}' - \boldsymbol{K}]\boldsymbol{y_c}. \tag{14.26}$$

The number of degrees of freedom for this sum of squares, denoted v_{LOF}, is computed by subtracting the pure error degrees of freedom from residual sum of squares' degrees of freedom, which is equal to $n - \rho$, that is,

$$v_{LOF} = n - \rho - \sum_{i=1}^{n_0}(v_i - 1). \tag{14.27}$$

In preparation for the development of the multivariate lack of fit test, we first denote the matrices $\boldsymbol{Y}'[\boldsymbol{I}_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-}\mathcal{X}' - \boldsymbol{K}]\boldsymbol{Y}$ and $\boldsymbol{Y}'\boldsymbol{K}\boldsymbol{Y}$ that are used in (14.26) and (14.22), respectively, by $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$, that is,

$$\boldsymbol{G}_1 = \boldsymbol{Y}'[\boldsymbol{I}_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-}\mathcal{X}' - \boldsymbol{K}]\boldsymbol{Y}, \tag{14.28}$$

$$\boldsymbol{G}_2 = \boldsymbol{Y}'\boldsymbol{K}\boldsymbol{Y}. \tag{14.29}$$

The matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ will be referred to as the lack of fit and pure error matrices, respectively.

Under the normality assumption made earlier concerning the distribution of the univariate model for $\boldsymbol{y_c}$ in (14.19), testing for its lack of fit can be carried out by using the ratio [see, e.g., Khuri and Cornell (1996), Section 2.6],

$$F(\boldsymbol{c}) = \frac{\nu_{PE}}{\nu_{LOF}} \frac{SS_{LOF}(\boldsymbol{c})}{SS_{PE}(\boldsymbol{c})}, \tag{14.30}$$

which can also be expressed as

$$F(\boldsymbol{c}) = \frac{\nu_{PE}}{\nu_{LOF}} \frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}. \tag{14.31}$$

If the model for $\boldsymbol{y_c}$ in (14.19) is correct, then $F(\boldsymbol{c})$ has the F-distribution with ν_{LOF} and ν_{PE}. A large value of $F(\boldsymbol{c})$, or equivalently, a large value of $\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}$ indicates that model (14.19) is inadequate for some $\boldsymbol{c} \neq 0$. Since the multireponse model in (14.18) is considered inadequate if at least one of the univariate models in (14.19) is inadequate for some $\boldsymbol{c} \neq 0$, we can then state that model (14.18) has a significant lack of fit if $\sup_{\boldsymbol{c}\neq 0} \frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}$ exceeds a certain critical value. To complete the development of this multivariate lack of fit test, the following lemma is needed.

Lemma 14.1

$$\sup_{\boldsymbol{c}\neq 0} \frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}} = e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1), \tag{14.32}$$

where $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1)$ *is the largest eigenvalue of the* $r \times r$ *matrix* $\boldsymbol{G}_2^{-1}\boldsymbol{G}_1$.

Proof. As shown in the proof of Corollary 9.3, the ratio $\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}$ can be written as

$$\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}} = \frac{\boldsymbol{w}'\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{w}}{\boldsymbol{w}'\boldsymbol{w}}, \tag{14.33}$$

where $\boldsymbol{w} = \boldsymbol{G}_2^{\frac{1}{2}}\boldsymbol{c}$. Using Corollary 9.2, we have

$$e_{min}\left(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\right) \leq \frac{\boldsymbol{w}'\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{w}}{\boldsymbol{w}'\boldsymbol{w}} \leq e_{max}\left(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\right), \tag{14.34}$$

where $e_{min}(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}})$ and $e_{max}(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}})$ are the smallest and the largest eigenvalues of $\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}$, respectively. Equalities in (14.34) hold for the lower and upper bounds if $\boldsymbol{w}$ is chosen an eigenvector of $\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}$ corresponding to $e_{min}(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}})$ and $e_{max}(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}})$, respectively. Making this choice for the upper bound in (14.34), we get

$$\sup_{\boldsymbol{w}\neq\boldsymbol{0}} \frac{\boldsymbol{w}'\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{w}}{\boldsymbol{w}'\boldsymbol{w}} = e_{max}\left(\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}\right). \tag{14.35}$$

Note that since the rank of $\boldsymbol{K}$ in (14.29) is ν_{PE}, the $r \times r$ matrix $\boldsymbol{G}_2$ will be positive definite, hence nonsingular, with probability 1 if $r \leq \nu_{PE}$ [see Roy et al. (1971, p. 35) and McDonald (1975, p. 463)]. We also note that the eigenvalues of the matrix $\boldsymbol{G}_2^{-\frac{1}{2}}\boldsymbol{G}_1\boldsymbol{G}_2^{-\frac{1}{2}}$ are real since it is symmetric by Lemma 6.1, and are also equal to the eigenvalues of $\boldsymbol{G}_2^{-1}\boldsymbol{G}_1$ by Theorem 6.12. We therefore conclude the validity of formula (14.32). ■

On the basis of Lemma 14.1 we can state that a significant lack of fit is detected at the α level if

$$e_{max}\left(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1\right) \geq \lambda_\alpha, \tag{14.36}$$

where λ_α is the upper $100\alpha\%$ point of the distribution of $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1)$ when the multiresponse model in (14.18) is correct. This multivariate lack of fit test is known as *Roy's largest-root test*. Tables for the values of λ_α are available in Roy et al. (1971) and Morrison (1976). Foster and Rees (1957) and Foster (1957) provide critical values for two and three responses, respectively. Heck (1960) presented some of such critical values in the form of charts.

Other multivariate lack of fit tests are available. These include *Wilks' likelihood ratio*, $|\boldsymbol{G}_2|/|\boldsymbol{G}_1 + \boldsymbol{G}_2|$, *Pillai's trace*, $tr[\boldsymbol{G}_1(\boldsymbol{G}_1 + \boldsymbol{G}_2)^{-1}]$, and *Lawley–Hotelling's trace*, $tr(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1)$. Small values of Wilks' likelihood ratio are significant, whereas large values of Pillai's trace and Lawley–Hotelling's trace are significant. Tables for the critical values of these tests can be found in Seber (1984). For a description of such tests, see, for example, Muirhead (1982, Chapter 10).

In the event of a significant lack of fit test, a procedure was proposed by Khuri (1985) for detecting which responses are responsible for lack of fit. This is similar to the problem of detecting differences among the treatment means in an analysis of variance whenever the F-test for testing equality of treatment means is significant. In our case, the comparisons are with respect to the responses rather than the means of the treatments.

TABLE 14.1 Design Settings and Multiresponse Values for Example 14.1

Design (original)		Design (coded)		Responses		
X_1 (mM)*	X_2 (mM)	x_1	x_2	y_1 (kg)	y_2 (mm)	y_3 (g)
21.0	16.2	0	0	1.50	1.80	0.44
21.0	16.2	0	0	1.66	1.79	0.50
21.0	16.2	0	0	1.48	1.79	0.50
21.0	16.2	0	0	1.41	1.77	0.43
21.0	16.2	0	0	1.58	1.73	0.47
08.0	06.5	−1	−1	2.48	1.95	0.22
34.0	06.5	1	−1	0.91	1.37	0.67
08.0	25.9	−1	1	0.71	1.74	0.57
34.0	25.9	1	1	0.41	1.20	0.69
02.6	16.2	−1.414	0	2.28	1.75	0.33
39.4	16.2	1.414	0	0.35	1.13	0.67
21.0	02.5	0	−1.414	2.14	1.68	0.42
21.0	29.9	0	1.414	0.78	1.51	0.57

* milliMolar

Source: A. I. Khuri and M. Conlon (1981). Reproduced with permission of the *American Statistical Association*.

Additional details concerning various aspects regarding the analysis of multiresponse experiments can be found in Khuri (1986), Khuri (1990 a), Khuri (1990 b), Khuri (1990 c), Khuri (1996), Khuri and Valeroso (1998), Khuri and Valeroso (2000), Valeroso and Khuri (1999).

Example 14.1 *Schmidt et al. (1979) investigated the effects of cysteine (X_1) and calcium chloride (X_2) on the textural and water-holding characteristics of dialyzed whey protein concentrates gel systems. Some of the characteristics measured included the following responses: hardness (y_1), springiness (y_2), and compressible water (y_3). A central composite design with five replications at the center was used. The design settings, in the original and coded settings, and the corresponding multiresponse data are shown in Table (14.1). This same data set was used in Khuri and Conlon(1981). The fitted model, in the coded variables, for each of the three responses is a second-degree model of the form*

$$y = \beta_0 + \sum_{i=1}^{2} \beta_i x_i + \beta_{12} x_1 x_2 + \sum_{i=1}^{2} \beta_{ii} x_i^2 + \epsilon.$$

In this case, the matrix $\mathcal{X} = [\boldsymbol{X}_1 : \boldsymbol{X}_2 : \boldsymbol{X}_3]$ in model (14.18) is of order 13×18 and rank $\rho = 6$ since $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \boldsymbol{X}_3$. We also have five replications at the center point, hence the matrix $\boldsymbol{K}$ in (14.23) is of the form $\boldsymbol{K} = diag(\boldsymbol{K}_1, \boldsymbol{0})$, where $\boldsymbol{K}_1 = \boldsymbol{I}_5 - \frac{1}{5}\boldsymbol{J}_5$. The pure error and lack of fit degrees of freedom are $\nu_{PE} = 4$ $\nu_{LOF} = 3$, respectively, since the number of degrees of freedom for the residual sum of squares is equal to $n - \rho = 7$. The matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are computed using (14.28) and (14.29), respectively, where $\boldsymbol{Y} = [\boldsymbol{y}_1 : \boldsymbol{y}_2 : \boldsymbol{y}_3]$. These matrices are given by

$$\boldsymbol{G}_1 = \begin{bmatrix} 0.2425133 & -0.044808 & -0.018044 \\ -0.044808 & 0.0143432 & -0.001655 \\ -0.018044 & -0.001655 & 0.0073032 \end{bmatrix} \qquad (14.37)$$

$$G_2 = \begin{bmatrix} 0.03712 & -0.00118 & 0.00806 \\ -0.00118 & 0.00312 & 0.00036 \\ 0.00806 & 0.00036 & 0.00428 \end{bmatrix} \tag{14.38}$$

The value of Roy's largest root test statistic in (14.36) is $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1) = 16.155$. *The critical value* λ_α *for this test can be obtained from tables of the upper percentage points of the* generalized Beta distribution *given in Foster (1957) for* $r = 3$ *responses. These tables actually give values of* x_α, *the upper* $100\alpha\%$ *point of the distribution of the largest eigenvalue of the matrix* $(\boldsymbol{G}_1 + \boldsymbol{G}_2)^{-1}\boldsymbol{G}_1$, *when model (14.18) is correct. The value of* λ_α *is related to* x_α *through the equation*

$$\lambda_\alpha = \frac{x_\alpha}{1 - x_\alpha}. \tag{14.39}$$

At the $\alpha = 0.10$ *level of significance,* $x_{0.10} \approx 0.9769$ *for 4 and 3 degrees of freedom for pure and lack of fit, respectively. Hence,* $\lambda_{0.10} = 42.29$. *We conclude that no significant lack of fit can be detected at the* 10% *level.*

REFERENCES

Box, G. E. P. and Draper, N. R. (1965). The Bayesian estimation of common parameters from several responses. *Biometrika*, 52, 355–365.

Box, G. E. P., Hunter, W. G., MacGregor, J. F., and Erjavec, J. (1973). Some problems associated with the analysis of multiresponse data. *Technometrics*, 15, 33–51.

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

Foster, F. G. (1957). Upper percentage points of the generalized Beta distribution II. *Biometrika*, 44, 441–453.

Foster, F. G. and Rees, D. H. (1957). Upper percentage points of the generalized Beta distribution I. *Biometrika*, 44, 237–247.

Heck, D. L. (1960). Charts of some upper percentage points of the distribution of the largest characteristic root. *Ann. Math. Stat.*, 31, 625–642.

Khuri, A. I. (1985). A test for lack of fit of a linear multiresponse model. *Technometrics*, 27, 213–218.

Khuri, A. I. (1986). Exact tests for the comparison of correlated response models with an unknown dispersion matrix. *Technometrics*, 28, 347–357.

Khuri, A. I. (1990a). The analysis of multiresponse experiments: A review. In: *Statistical Design and Analysis of Industrial Experiments*, pp. 231–246, edited by S. Ghosh. Dekker, New York.

Khuri, A. I. (1990b). Multiresponse rotatability. *J. Stat. Plan. Inf.*, 25, 1–6.

Khuri, A. I. (1990c). The effect of response scaling on the detection of linear dependencies among multiresponse data. *Metrika*, 37, 217–231.

Khuri, A. I. (1996). Multiresponse surface methodology. In: *Handbook of Statistics*, vol. 13, pp. 377–406, edited by S. Ghosh and C. R. Rao. Elsevier Science B. V., Amsterdam.

Khuri, A. I. and Conlon, M. (1981). Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, 23, 363–375.

Khuri, A. I. and Cornell, J. A. (1996). *Response Surfaces*, 2nd ed. Dekker, New York.

Khuri, A. I. and Valeroso, E. S. (1998). Multiresponse surface models with block effects. *J. Stat. Plan. Inf.*, 73, 7–20.

Khuri, A. I. and Valeroso, E. S. (2000). Optimization methods in multiresponse surface methodology. In: *Statistical Process Monitoring and Optimization*, pp. 411–433, edited by S. H. Park and G. G. Vining. Dekker, New York.

McDonald, L. (1975). Tests for the general linear hypothesis under the multiple design multivariate linear model. *Ann. Stat.*, 3, 461–466.

Morrison, D. F. (1976). *Multivariate Statistical Methods*, 2nd ed. McGraw-Hill, New York.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Stat.*, 24, 220–238.

Roy, S. N., Gnanadesikan, R., and Srivastava, J. N. (1971). *Analysis and Design of Certain Quantitative Multiresponse Experiments*. Pergamon Press, New York.

Schmidt, R. H., Illingsworth, B. L., Deng, J. C., and Cornell, J. A. (1979). Multiple regression and response surface analysis of the effects of calcium chloride and cysteine on heat-induced whey protein gelation. *J. Agric. Food Chem.*, 27, 529–532.

Seber, G. A. F. (1984). *Multivariate Observations*. John Wiley & Sons, New York.

Valeroso, E. S. and Khuri, A. I. (1999). Multiresponse surface models with random block effects. *J. Stat. Plan. Inf.*, 79, 157–173.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Stat. Assoc.*, 57, 348–368.

EXERCISES

14.1 Consider the matrix $\boldsymbol{G}$ given in (14.8). Show that m linearly independent relationships exist among the r responses, of the type shown in (14.5), if and only if $\boldsymbol{GG}'$ has a zero eigenvalue of multiplicity m.

14.2 Show that if $\boldsymbol{GG}'$ has a zero eigenvalue of multiplicity m, then the corresponding eigenvectors of $\boldsymbol{GG}'$ define m linearly independent relationships among the responses.

14.3 Consider the multiresponse values in Table 14.1 for Example 14.1. Determine if the matrix $\boldsymbol{GG}'$ has any zero eigenvalues. If so, what are the corresponding linear relationships among the three responses?

Note: The SAS statements needed for this exercise are the following:
PROC IML;
$yy_1 = \{1.50\ 1.66\ 1.48\ 1.41\ 1.58\ 2.48\ 0.91\ 0.71\ 0.41\ 2.28\ 0.35\ 2.14\ 0.78\}$;
$y_1 = y\grave{y}_1$;
$yy_2 = \{1.80\ 1.79\ 1.79\ 1.77\ 1.73\ 1.95\ 1.37\ 1.74\ 1.20\ 1.75\ 1.13\ 1.68\ 1.51\}$;
$y_2 = y\grave{y}_2$;
$yy_3 = \{0.44\ 0.50\ 0.50\ 0.43\ 0.47\ 0.22\ 0.67\ 0.57\ 0.69\ 0.33\ 0.67\ 0.42\ 0.57\}$;
$y_3 = y\grave{y}_3$;
$\boldsymbol{Y} = \boldsymbol{y_1} \,||\, \boldsymbol{y_2} \,||\, \boldsymbol{y_3}$;
PRINT **Y**;
$\boldsymbol{G} = \grave{\boldsymbol{Y}} * (\boldsymbol{I}(13) - (1/13) * \boldsymbol{J}(13, 13, 1))$;
$\boldsymbol{H} = \boldsymbol{G} * \grave{\boldsymbol{G}}$;
CALL EIGEN(**M**, **E**, **H**);

PRINT **M**;

y_1	y_2	y_3	y_4
86.5	7.9	2.0	0.3
75.3	16.1	4.8	0.6
66.1	22.9	5.1	1.2
51.9	33.6	5.3	1.4
38.9	41.8	6.8	1.8
25.4	48.8	6.1	2.3
13.8	56.3	5.8	2.5
4.9	62.8	3.7	2.8

PRINT **E**;

14.4 Consider the multiresponse data given in the table above. Determine if there are linear relationships among the four responses.

14.5 Show that if the $\boldsymbol{X}_i$'s are all equal to $\boldsymbol{X}_0$ in (14.13), then $\hat{\boldsymbol{\beta}}$ takes the form

$$\hat{\boldsymbol{\beta}} = \left[\boldsymbol{I}_r \otimes (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\right]\boldsymbol{y}.$$

14.6 Suppose that model (14.19) is inadequate for some $\boldsymbol{c} \neq \boldsymbol{0}$. Show that an eigenvector of $\boldsymbol{G}_2^{-1}\boldsymbol{G}_1$ corresponding to its largest eigenvalue, $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1)$, is such a vector, where $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are given in (14.28) and (14.29), respectively.
(Hint: Find the vector $\boldsymbol{c}$ which maximizes $\dfrac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}$ by taking its derivative with respect to $\boldsymbol{c}$ and equating it to zero.)

14.7 Consider the following data from a multiresponse experiment:

x_1	x_2	x_3	x_4	y_1	y_2	y_3
0	0	0	0	64.01	45.88	21.03
0	0	0	0	65.25	47.77	28.33
0	0	0	0	61.01	41.04	29.5
0	0	0	0	60.58	49.18	30.95
0	0	0	0	66.66	47.09	30.27
−1	−1	−1	−1	57.92	53.95	38.94
1	−1	−1	−1	77.32	55.36	45.7
−1	1	−1	−1	48.72	50.71	23.56
1	1	−1	−1	70.65	38.01	29.3
−1	−1	1	−1	48.39	59.92	25.41
1	−1	1	−1	68.61	56.91	30.72
−1	1	1	−1	35.19	59.02	29.01
1	1	1	−1	59.28	48.70	25.67
−1	−1	−1	1	74.66	46.82	27.95
1	−1	−1	1	80.05	31.41	29.93
−1	1	−1	1	61.01	38.77	28.70
1	1	−1	1	79.77	31.95	26.19
−1	−1	1	1	66.12	66.86	36.66
1	−1	1	1	70.11	46.30	30.20
−1	1	1	1	52.19	62.86	38.09
1	1	1	1	64.66	40.65	29.75

The fitted response models in x_1, x_2, x_3, and x_4 are

$$y_1 = \beta_{10} + \sum_{i=1}^{4} \beta_{1i}x_i + \beta_{114}x_1x_4$$

$$y2 = \beta_{20} + \sum_{i=1}^{4} \beta_{2i}x_i + \beta_{234}x_3x_4$$

$$y_3 = \beta_{30} + \sum_{i=2}^{4} \beta_{3i}x_i + \beta_{324}x_2x_4 + \beta_{334}x_3x_4.$$

Test for lack of fit of the fitted models using Roy's largest root test. Let $\alpha = 0.10$.

14.8 Consider again Exercise 14.7. Suppose that lack of fit is significant. Identify a linear combination of the responses for which model (14.19) is inadequate.
(Hint: Use the result in Exercise 14.6.)

Part III

Matrix Computations and Related Software

This part presents an overview of the software available for doing matrix computations for the benefit of the reader. The intention here is to make it easy for the reader to do most of the matrix computations described in the book. The following three chapters are included in this part:

Chapter 15: SAS/IML

SAS/IML (*Interactive Matrix Language*) is an interactive programming language that can be used for performing matrix computations using functions and algorithms associated with matrices. It is part of the more general SAS software system for statistical data analysis, where SAS stands for *Statistical Analysis System*. IML is a powerful system that is easy to use and is quite suited for carrying out statistical procedures involving linear models. It can easily handle many aspects of matrix algebra such as matrix decomposition and eigenvalue problems. It also allows the user to have access to built-in operators and call routines to do complex matrix operations. Furthermore, IML has the capability to perform a wide variety of statistical procedures such as regression analysis, robust regression, and mixed model analysis, to name just a few. A set of graphics commands are also available from which to generate plots and customized displays. IML can therefore provide the reader with a lot more than is needed to do the matrix manipulations in this book.

SAS/IML is a product of the SAS Institute, Inc., Cary, North Carolina, USA. The current version of SAS/IML used in this book is Version 9.4.

The SAS statements used earlier in this book will be revisited for further clarifications. In fact, after going through this chapter, the reader can easily understand and follow the choice of such statements.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

Chapter 16: Use of MATLAB in Matrix Computations

This chapter covers the use of MATLAB in matrix computations which includes an introduction to mathematical functions, arithmetic operators, and construction of matrices. Two- and three-dimensional plots are also discussed.

Chapter 17: Use of R in Matrix Computations

This chapter includes construction of matrices, various R commands and functions for carrying out certain operations such as the Cholesky decomposition, spectral decomposition, and singular-value decomposition. Two- and three-dimensional plots are described.

15

SAS/IML

15.1 GETTING STARTED

SAS/IML has three windows: the *Editor* window, the *Output* window, and the *Log* window. The Editor window contains the SAS statements entered after opening a new program or a program that was previously saved. After submitting the program the results of the execution appear in the Output window. Any error messages are displayed in the Log window which can be used to fix any problems that prevent showing all the intended results.

SAS is not case-sensitive as uppercase and lowercase letters can be used. For better clarity, all SAS statements will be described here using uppercase letters.

To begin a SAS session, get into the SAS system and enter

PROC IML; (Note that all SAS statements end in a semicolon)

15.2 DEFINING A MATRIX

Numeric as well as character matrices can be used. For a numeric matrix, curly brackets are used and the rows of a matrix are entered separated by commas. For example, we define the 3×3 matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \{4 \;\; 9 \;\; 14, 6 \;\; 16 \;\; 10, 12 \;\; 23 \;\; 8\},$$

which gives

$$\boldsymbol{A} = \begin{bmatrix} 4 & 9 & 14 \\ 6 & 16 & 10 \\ 12 & 23 & 8 \end{bmatrix}. \tag{15.1}$$

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

The column vector $\boldsymbol{a} = (-3, 6, 10, 9)'$ is written as

$$\boldsymbol{a} = \{-3, 6, 10, 9\},$$

which gives

$$\boldsymbol{a} = \begin{bmatrix} -3 \\ 6 \\ 10 \\ 9 \end{bmatrix}.$$

An example of a character matrix is one giving the ages of three individuals:

$$\mathbf{Age} = \{John\ \ 25,\ Alan\ \ 29,\ Jim\ \ 32\},$$

which gives

$$\mathbf{Age} = \begin{bmatrix} John & 25 \\ Alan & 29 \\ Jim & 32 \end{bmatrix}.$$

To point to a particular element in a matrix, for example, the element in row 2 and column 3 in the matrix $\boldsymbol{A}$ in (15.1) we write $b = \boldsymbol{A}[2, 3]$; which has the value 10. To refer to a row or column in $\boldsymbol{A}$, we write $\boldsymbol{r} = \boldsymbol{A}[2,]$; which assigns the second row of $\boldsymbol{A}$ to $\boldsymbol{r}$. The statement $\boldsymbol{c} = \boldsymbol{A}[, 3]$; assigns the third column of $\boldsymbol{A}$ to $\boldsymbol{c}$. In addition, the statement $\boldsymbol{B} = \boldsymbol{A}[+,]$; assigns to $\mathbf{B}$ the column sums of $\boldsymbol{A}$, and $\boldsymbol{C} = \boldsymbol{A}[, +]$; assigns to $\mathbf{C}$ the row sums of $\boldsymbol{A}$. Furthermore, $\boldsymbol{F} = \boldsymbol{A}[\{2\ \ 3\}, +]$; produces the sums of rows 2 and 3 of $\boldsymbol{A}$, and $E = \boldsymbol{A}[:]$; gives the mean of all elements of $\boldsymbol{A}$.

To select a submatrix of $\boldsymbol{A}$, for example, the 2×2 submatrix of $\boldsymbol{A}$ consisting of the four elements in the lower-left corner of $\boldsymbol{A}$, we write $\boldsymbol{D} = A[\{2\ \ 3\}, \{1\ \ 2\}]$; which gives the matrix

$$\boldsymbol{D} = \begin{bmatrix} 6 & 16 \\ 12 & 23 \end{bmatrix}.$$

The statement $\boldsymbol{v} = \boldsymbol{A}[1 : 3, 1]$; is equivalent to $\boldsymbol{v} = \boldsymbol{A}[\{1\ \ 2\ \ 3\}, 1]$; which defines $\boldsymbol{v}$ to be the 3×1 matrix consisting of the three elements in the first column of $\boldsymbol{A}$.

15.3 CREATING A MATRIX

A matrix can be created by manual entry as described in the previous section. Its contents can be viewed by using the statement “print A;”. A matrix can also be created from a SAS data set. For example, consider the data set of Example 13.6 which is given in Table 13.4. It concerns a two-way crossed classification experiment with two factors denoted by A and B. We would like to create a vector consisting of the 18 values of the response y shown in Table 13.4. This is demonstrated by first creating the data set using the following SAS statements:

```
DATA ONE;
INPUT A B y@@;
CARDS;
1 1 72 1 1 60 1 1 62 1 2 75 1 2 71 1 2 72 1 3 92 1 3 101 1 3 91
2 1 101 2 1 96 2 1 109 2 2 110 2 2 108 2 2 101 2 3 116 2 3 110 2 3 106
DATA TWO;
SET ONE;
RUN;
```

Note that a "RUN" statement is needed at the end of a general SAS program, but is not needed in PROC IML. The next step is to use PROC IML to create the vector of 18 response values. This is done as follows:

```
PROC IML;
USE TWO;
READ ALL INTO G;
y=G[1:18,3];
PRINT y;
```

Note that the data set named TWO consists of the levels of A and B in addition to the y values. Thus the matrix matrix **G** should have three columns; the first two give the levels of A and B and the third column consists of the y values. Hence, **G**[1:18,3] gives the desired vector of response values.

15.4 MATRIX OPERATIONS

We now provide a listing of operators and functions used in IML. This listing is based on Chapter 20 in SAS (2004).

Operation

Addition + A+B

Subtraction - A-B

Matrix multiplication * **A*B**

Element multiplication # **A#B**

Note: This operator produces a matrix whose elements are the products of the corresponding elements of **A** and **B**.

Direct product @ **A@B**

Division / **A/B**

This operator divides each element of **A** by the corresponding elements of **B**. It can also be used to divide a matrix by a scalar.

Matrix power ** **A****scalar

This operator creates a matrix that is **A** multiplied by itself scalar times. The matrix must be square and the scalar must be an integer greater than or equal to -1. For example, if **A** is the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix},$$

then $\boldsymbol{A} ** 2$ gives the matrix $\boldsymbol{A} * \boldsymbol{A}$, that is,

$$\boldsymbol{A} ** 2 = \begin{bmatrix} 19 & 9 \\ 15 & 16 \end{bmatrix}.$$

Element power ## **A##B**

This operator creates a matrix whose elements are the elements of **A** raised to the power from the corresponding element of **B**. For example, if $\boldsymbol{A} = \{4 \;\; 2 \;\; 3 \;\; 5\}$, then $\boldsymbol{A}$##2 is equal to $\{16 \;\; 4 \;\; 9 \;\; 25\}$.

Element maximum <> **A<>B**

This operator compares each element of **A** to the corresponding element of **B**. The larger of the two values becomes the corresponding element of the new matrix. The two matrices must have the same dimension. For example, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are given by

$$\boldsymbol{A} = \begin{bmatrix} 3 & 6 & 7 \\ 9 & 10 & 14 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} 2 & 12 & 1 \\ 15 & 8 & 30 \end{bmatrix}$$

then

$$\boldsymbol{A} <> \boldsymbol{B} = \begin{bmatrix} 3 & 12 & 7 \\ 15 & 10 & 30 \end{bmatrix}.$$

Element minimum >< **A><B**

This is similar to the element maximum operator, except that the smaller of the two values becomes the corresponding element of the new matrix.

Horizontal concatenation || **A||B**

This operator produces a new matrix by horizontally joining **A** to **B**. The two matrices must have the same number of rows. More than two matrices can also be joined horizontally using this operator.

Vertical concatenation // **A//B**

This operator produces a new matrix by joining **A** and **B** vertically. The two matrices must have the same number of columns.

Transpose operator ‘ **À**

This operator exchanges the rows and columns of **A**. Thus if **A** is an $m \times n$ matrix, then **À** is of order $n \times m$.

ABS function ABS(**A**)

This function gives the absolute value of every element of the matrix $\boldsymbol{A}$.

APPEND APPEND FROM **A**

This statement adds observations from matrix **A** to a SAS data set.

BLOCK function BLOCK(**A**,**B**)

This function creates a block-diagonal matrix from all the matrices listed. Thus Block(**A**,**B**) is a matrix of the form

$$BLOCK(\boldsymbol{A}, \boldsymbol{B}) = \begin{bmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{B} \end{bmatrix}.$$

CLOSE CLOSE (SAS data set) This statement closes a SAS data set.

CREATE CREATE SAS data set FROM A

This statement creates a new data set from matrix A

DET function DET(**A**)

This function computes the determinant of the square matrix **A**.

DIAG function DIAG(**A**)

This function creates a matrix whose diagonal elements are equal to the corresponding diagonal elements of **A**. The off-diagonal elements are all equal to zero. For example, if **A** is the matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 & -1 \\ 8 & 2 & -9 \\ 6 & 0 & 11 \end{bmatrix},$$

then DIAG(**A**) is expressed as

$$DIAG(\boldsymbol{A}) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 11 \end{bmatrix}.$$

DO;

statements

END;

The DO statement specifies that the statements after DO are executed as a group until the END statement appears.

EIGEN call: eigenvalues and eigenvectors. CALL EIGEN(**M,E,A**)

The Eigen subroutine creates eigenvalues of a symmetric matrix **A** of order $n \times n$. The eigenvalues are the elements of a matrix **M** arranged in descending order. The matrix **E** gives the eigenvectors of **A** which correspond to its eigenvalues. The matrix **E** is orthogonal whose columns are orthonormal. The first column of **E** corresponds to the first eigenvalue in **M**, and so forth. In this case, the matrix **A** can be written as $\boldsymbol{A} = \boldsymbol{E} * diag(\boldsymbol{M}) * \grave{\boldsymbol{E}}$ as described in the *spectral decomposition theorem* (see Theorem 6.2).

EIGVAL function EIGVAL(**A**)

This function gives the eigenvalues of a symmetric matrix **A**. The eigenvalues are real and are presented as a column vector.

EIGVEC function EIGVEC(**A**)

This function creates the eigenvectors of a symmetric matrix **A**.

END statement

This statement ends a DO loop or DO statement.

EXP function EXP(**A**)

This function takes the exponential function of every element of the matrix **A**.

FINISH statement

This statement denotes the end of a module.

GINV function GINV(**A**)

This function gives the Moore–Penrose generalized inverse of the matrix **A**.

ROOT function ROOT(**A**)

This function gives the Cholesky decomposition of the matrix A such that

$$A = \grave{U}U,$$

where U is upper triangular. The matrix A must be symmetric and non-negative definite.

IF (expression) THEN statement1;
ELSE statement2;
The IF statement contains an expression to be evaluated. If the expression is true, then the statement1 is executed. If the expression is false, then statement2 is executed. For example, we have
If $A > 1$ Then b=2;
ELSE b=0;

I function **I**(n)

This function gives the identity matrix of size n.

INV function INV(**A**)

This function gives the inverse of the matrix A, which is assumed to be square and nonsingular.

J function **J**(m,n,v)

This function creates a matrix with m rows, n columns, with all elements equal to the value v.

LOG function lOG(**A**)

This function takes the natural logarithm of each element of the matrix **A**.

NCOL function NCOL(**A**)

This function gives the number of columns of the matrix **A**.

NORMAL function NORMAL(seed)

This function gives a random number having a normal distribution with mean 0 and a standard deviation of 1. The normal function returns a matrix with the same dimension as the argument.

NROW function NROW(**A**)

This function gives the number of rows of the matrix **A**.

PRINT statement PRINT **A**

This statement is used to print a specified matrix such as **A**.

PROBCHI function PROBCHI(x,df,nc)

This function gives the probability that an observation from a chi-squared distribution, with df degrees of freedom and a noncentrality parameter nc, is less than or equal to x. If nc is not specified or has the value zero, the probability for a central chi-squared distribution is given.

PROBF function PROBF(x,ndf,ddf,nc)

This function gives the probability that an observation from an F-distribution, with ndf numerator degrees of freedom, ddf denominator degrees of freedom, and a noncentrality parameter nc, is less than or equal to x.

PROBNORM function PROBNORM(x)

This function gives the probability that an observation from the standard normal distribution is less than or equal to x.

PROBT function PROBT(x, df, nc)
This function gives the probability that an observation from the t distribution, with df degrees of freedom and a noncentrality parameter nc, is less than or equal to x.

RANK function RANK(**A**)
This function produces a matrix whose elements are the ranks of the corresponding elements of the matrix **A**.

READ ALL INTO A
This statement is used to read variables from a SAS data set into columns of the matrix $\boldsymbol{A}$.

RUN statement
This statement executes a module.

SQRT function SQRT(**A**)
This function computes the positive square roots of each element of the matrix $\boldsymbol{A}$.

SSQ function SSQ(**A**)
This function computes the sum of squares of all the elements of the matrix $\boldsymbol{A}$.

START statement
This statement defines a module.

SUM function SUM(**A**)
This function gives the sum of all elements in the matrix **A**.

STOP statement STOP
This statement stops the IML program.

SVD call call SVD(**U**,**Q**,**V**,**A**)
This subroutine produces the singular-value decomposition (see Section 7.4) of an $m \times n$ matrix $\boldsymbol{A}$ of rank r, where $\boldsymbol{U}$, $\boldsymbol{Q}$, $\boldsymbol{V}$ are decomposition matrices.

TRACE function TRACE(**A**)
This function computes the sum of the diagonal elements of the square matrix $\boldsymbol{A}$.

UNIFORM function UNIFORM(seed)
This function produces a random number having the uniform distribution over the interval [0,1]. The function returns a matrix with the same dimension as the argument. The first argument on the first call is used for the seed.

USE statement SAS data set
This statement is used to open a SAS data set for reading, for example, USE ONE; specifies a data set called ONE.

VECDIAG function VECDIAG(**A**)
This function gives a column vector consisting of the diagonal elements of the square matrix $\boldsymbol{A}$.

Note that a module is a user-defined function or subroutine that can be called from an IML program. Modules are used for creating groups of statements which can be invoked as a unit from anywhere in the SAS program.

Example 15.1 *Consider the following matrix:*

$$\boldsymbol{A} = \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 4 \end{bmatrix}$$

The following items that pertain to **A** *are computed using PROC IML:*

1. *The determinant of* **A** *is det(***A***)* = −*5.*
2. *The inverse of* **A** *is*

$$\mathbf{A}^{-1} = \begin{bmatrix} -1.6 & 2.4 & -0.6 \\ -0.2 & 0.8 & -0.2 \\ 0.4 & -0.6 & 0.4 \end{bmatrix}$$

3. *The trace of* **A** *is trace(***A***)* = *5.*
4. *The eigenvalues of* **A** *are given by the elements of the column vector,*

Eigenvalues
4.253418
1.5199694
−0.773387

The corresponding eigenvectors are given by the column of the following matrix with the first column corresponding to the first eigenvalue and so forth:

Eigenvectors		
−0.22566	0.7316305	0.9760888
−0.395161	0.6145622	0.0737313
−0.890464	−0.295009	−0.204486

5. *The sum of the elements of* **A** *is sum(***A***)* = *10.*
6. *The sum of squares of the elements of* **A** *is ssq(***A***)* = *32.*
7. *The ranks of the elements of* **A** *are given by the following matrix:*

$$rank(\mathbf{A}) = \begin{bmatrix} 1 & 8 & 2 \\ 3 & 7 & 5 \\ 6 & 4 & 9 \end{bmatrix}.$$

The SAS statements used in this example are
PROC IML;
A = *{-1 3 0,0 2 1,1 0 4};*
*d=DET(***A***); PRINT d;*
*I=INV(***A***); PRINT I;*
*t=TRACE(***A***); PRINT t;*
*CALL EIGEN(***M***,***E***,***A***);*
PRINT **M***; PRINT* **E***;*
*s=SUM(***A***); PRINT s;*
*ss=SSQ(***A***); PRINT ss;*
*r=RANK(***A***); PRINT r;*

Example 15.2 *Consider the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

To verify that it is positive definite we need to show that its eigenvalues are positive. These eigenvalues are 6.4114741, 1.8152075, 0.7733184, hence $\boldsymbol{A}$ *is positive definite. Let us now find the Cholesky decomposition of* $\boldsymbol{A}$ *such that* $\boldsymbol{A} = \grave{\boldsymbol{U}}\boldsymbol{U}$*, where* $\boldsymbol{U}$ *is an upper triangular matrix. This matrix is*

$$\boldsymbol{U} = \begin{bmatrix} 1.4142136 & 1.4142136 & 0.7071068 \\ 0 & 1.7320508 & 0 \\ 0 & 0 & 1.2247449 \end{bmatrix}$$

The SAS statements used in this example are

```
PROC IML;
A={2 2 1,2 5 1,1 1 2};
CALL EIGEN(M,E,A);
PRINT M;
U=ROOT(A);
PRINT U;
```

Example 15.3 *Consider a data set which consists of ten values of x and corresponding response values:*

x	y
265	18
285	19
225	24
426	19
445	31
485	25
750	50
800	64
924	68
1108	71

A simple linear regression model was fitted using this data set,

$$y = \beta_0 + \beta_1 x + \epsilon.$$

The least-squares estimates of β_0 *and* β_1 *are* $\hat{\beta}_0 = -0.394721$*,* $\hat{\beta}_1 = 0.0687812$*, respectively. The residual sum of squares is* $SS_E = 377.62497$*. The estimated variance-covariance* *matrix of* $\hat{\boldsymbol{\beta}}$ *is*

$$\begin{aligned}\hat{var}(\hat{\boldsymbol{\beta}}) &= MS_E(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \begin{bmatrix} 23.129027 & -0.032223 \\ -0.032223 & 0.0000564 \end{bmatrix},\end{aligned}$$

where $MS_E = SS_E/8$ *and* $\hat{\boldsymbol{\beta}}$ *is the vector whose elements are* $\hat{\beta}_0$ *and* $\hat{\beta}_1$.

The SAS statements needed to get the results in this example are

```
PROC IML;
X={1 265,1 285,1 225,1 426,1 445,1 485,1 750,1 800,1 924,1 1108};
PRINT X;
y={18,19,24,19,31,25,50,64,68,71};
PRINT y;
XPI=INV(X`X);
BETAHAT=XPI*X`*y;
PRINT BETAHAT;
H=X*XPI*X`;
RESID=(I(10) - H)*y;
SSE=RESID`*RESID;
PRINT RESID;
PRINT SSE;
MSE=SSE/8;
PRINT MSE;
VCBETA=MSE*XPI;
PRINT VCBETA;
```

Example 15.4 *Let us again consider the data set of Example 15.3. It is assumed that the y values form a sample from a normal distribution with variance* σ^2, *hence the error term in the model has the same distribution and variance, but with mean equal to zero. Suppose that it is of interest to obtain 95% confidence intervals on* β_0, β_1, *and* σ^2. *The 95% confidence intervals on* β_0 *and* β_1 *are*

$$\hat{\beta}_0 \pm t_{0.025,n-2}\left[MS_E\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)\right]^{1/2},$$

$$\hat{\beta}_1 \pm t_{0.025,n-2}\left[\frac{MS_E}{Sxx}\right]^{1/2},$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $n = 10$, *and*

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2.$$

Using SAS, the numerical values of the confidence intervals are $(-11.48489, 10.695443)$ *for* β_0 *and* $(0.0514629, 0.0860996)$ *for* β_1.

The 95% confidence interval on σ^2 is

$$\left(\frac{SS_E}{\chi^2_{0.025,n-2}}, \ \frac{SS_E}{\chi^2_{0.975,n-2}}\right).$$

The corresponding numerical value of this interval is (21.541641, 173.22246).

The SAS statements used for this example are

PROC IML;

$\boldsymbol{X} = \{1\ 265, 1\ 285, 1\ 225, 1\ 426, 1\ 445, 1\ 485, 1\ 750, 1\ 800, 1\ 924, 1\ 1108\}$*;*

$\boldsymbol{y} = \{18, 19, 24, 19, 31, 25, 50, 64, 68, 71\}$*;*

$\boldsymbol{XPI} = INV(\grave{\boldsymbol{X}} * X)$*;*

$\boldsymbol{BETAHAT} = \boldsymbol{XPI} * \grave{\boldsymbol{X}} * \boldsymbol{y}$*;*

$\boldsymbol{BETAHAT}_0 = \boldsymbol{BETAHAT}[1, 1]$*;*

$\boldsymbol{BETAHAT}_1 = \boldsymbol{BETAHAT}[2, 1]$*;*

$\boldsymbol{H} = \boldsymbol{X} * \boldsymbol{XPI} * \grave{\boldsymbol{X}}$*;*

$\boldsymbol{RESID} = (\boldsymbol{I}(10) - \boldsymbol{H}) * \boldsymbol{y}$*;*

$SSE = \boldsymbol{RESID'} * \boldsymbol{RESID}$*;*

$MSE = SSE/8$*;*

$\boldsymbol{XVALUE} = \boldsymbol{X}[, 2]$ */* **XVALUE** is the vector of x values, which is the second column of X */;*

$SQUAREX = \boldsymbol{XVAL\grave{U}E} * \boldsymbol{XVALUE}$ */* SQUAREX is the sum of squares of the x values */;*

$SUMX = \boldsymbol{XVAL\grave{U}E} * J(10, 1, 1)$ */* SUMX is the sum of the x values */;*

$XBAR = SUMX/10$*;*

$SXX = SQUAREX - (SUMX ** 2)/10$*;*

$/ * t_{0.025,8} = 2.306 * /$*;*

$LCIBETA1 = BETAHAT_1 - 2.306 * SQRT(MSE/SXX)$ */* this is the lower 95% confidence limit on $BETA_1$ */;*

$UCBETA1 = BETAHAT_1 + 2.306 * SQRT(MSE/SXX)$ */* this is the upper 95% confidence limit on $BETA_1$ */;*

$LCIBETA0 = BETAHAT_0 - 2.306 * SQRT(MSE * (1/10 + (XBAR ** 2)/SXX))$ */* this is the lower 95% confidence limit on $BETA_0$ */;*

$UCIBETA0 = BETAHAT_0 + 2.306 * SQRT(MSE * (1/10 + (XBAR ** 2)/SXX))$ */* this is the upper 95% confidence limit on $BETA_0$ */;*

$/ * chisq_{0.025,8} = 17.53\ chisq_{0.975,8} = 2.18 * /$*;*

$LCISIGMASQ = SSE/17.53$ */* this is the lower 95% confidene limit on the error variance */;*

$UCISIGMASQ = SSE/2.18$ */* this is the upper 95% confidence limit on the error variance */;*

Example 15.5 *Consider again Example 15.4. It is of interest here to plot the residuals against x as well as the predicted response values against x. These plots are shown in Figures 15.1 and 15.2, respectively.*

The SAS statements used to generate Figure 15.1 are

ODS PDF FILE = 'PLOT _ of _ RESIDUALS _ VERSUS _ x . pdf';

PROC IML;

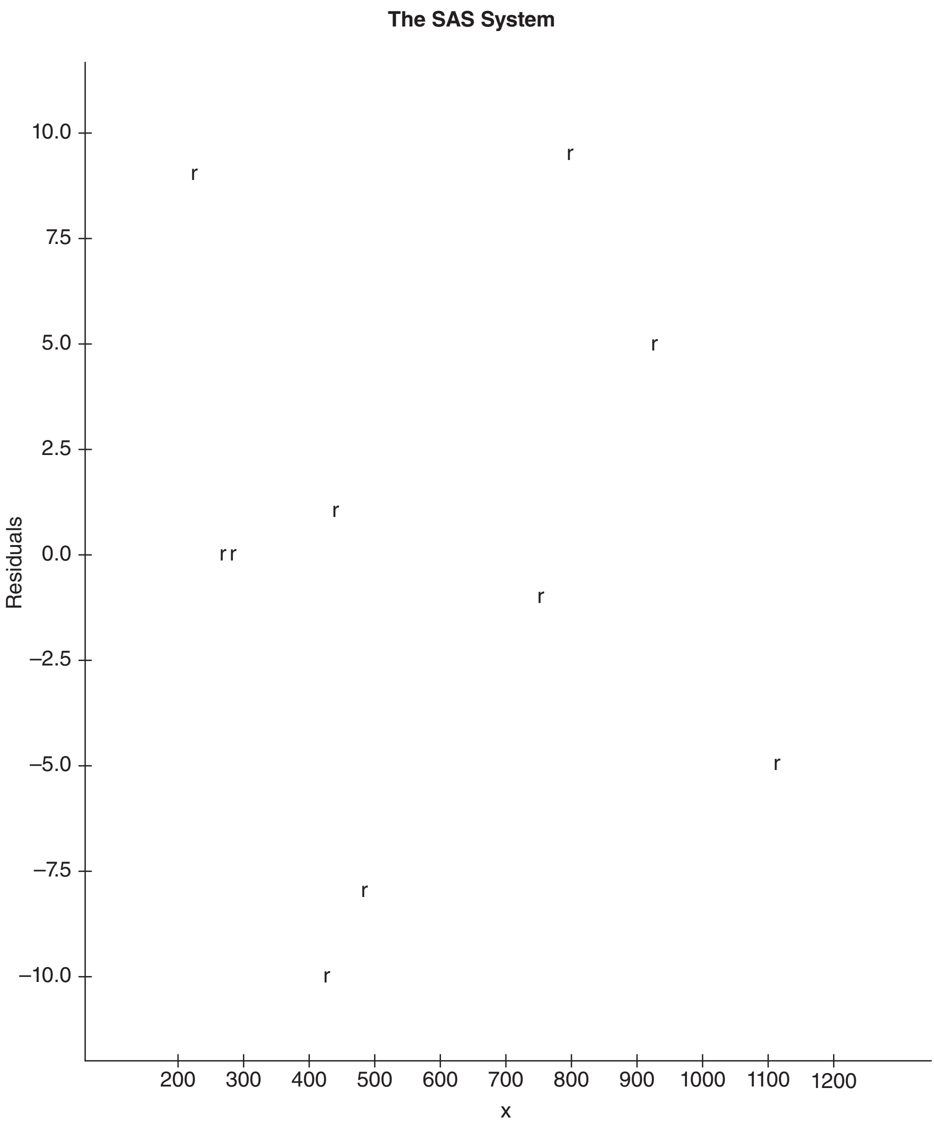


Figure 15.1 *Plot of Residuals Versus x.*

$\boldsymbol{X} = \{1\,265, 1\,285, 1\,225, 1\,426, 1\,445, 1\,485, 1\,750, 1\,800, 1\,924, 1\,1108\}$;
$\boldsymbol{y} = \{18, 19, 24, 19, 31, 25, 50, 64, 68, 71\}$;
$\boldsymbol{X1} = \boldsymbol{X}[, 2]$;
$\boldsymbol{XPI} = INV(\grave{\boldsymbol{X}} * \boldsymbol{X})$;
$\boldsymbol{BETA} = \boldsymbol{XPI} * \grave{\boldsymbol{X}} * \boldsymbol{y}$;
$\boldsymbol{H} = \boldsymbol{X} * \boldsymbol{XPI} * \grave{\boldsymbol{X}}$;
$\boldsymbol{RESID} = (\boldsymbol{I}(10) - \boldsymbol{H}) * \boldsymbol{y}$;
$\boldsymbol{XR} = \boldsymbol{X1}||\boldsymbol{RESID}$;
*CALL PGRAF(**XR**,'r','x','RESIDUALS','PLOT of RESIDUALS');*
ODS PDF CLOSE;

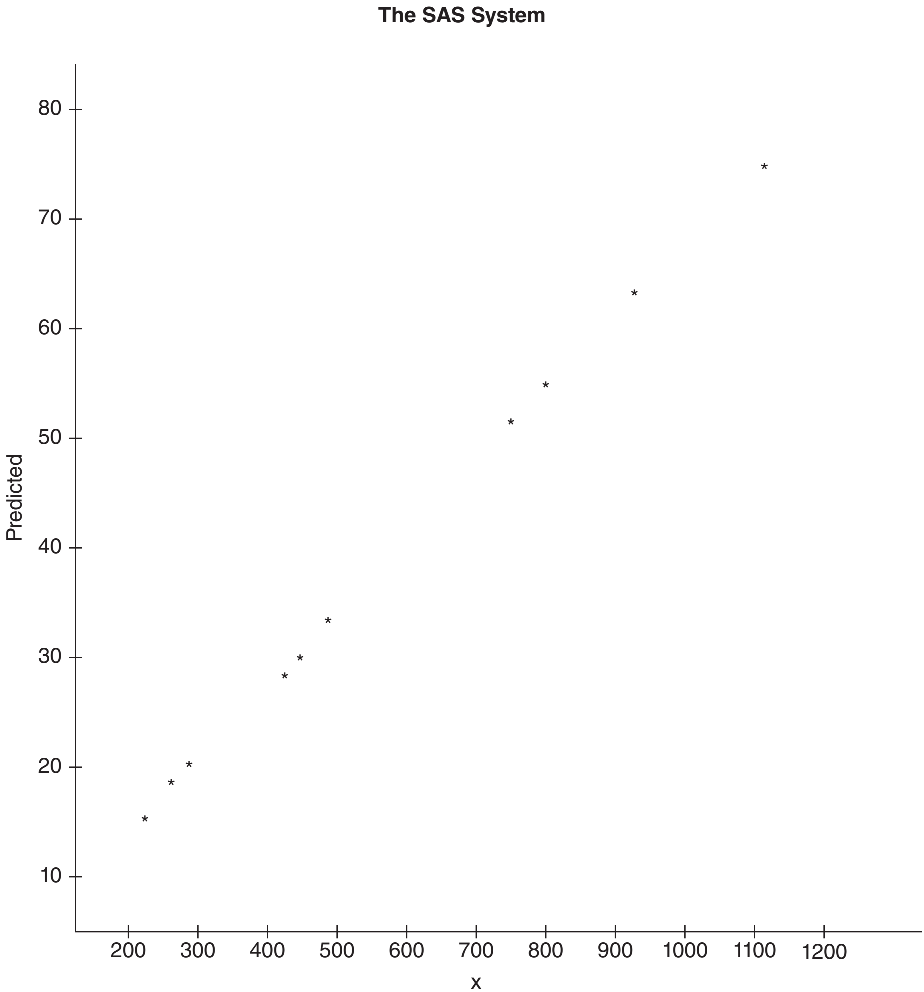


Figure 15.2 *Plot of Predicted Versus x.*

Note that the first and last statements are needed to store the SAS output in a PDF format. The SAS statements used for Figure 15.2 are

ODS PDF FILE = 'PLOT _ of _ PREDICTED _ VERSUS _ x . pdf';

$\boldsymbol{X} = \{1\ 265, 1\ 285, 1\ 225, 1\ 426, 1\ 445, 1\ 485, 1\ 750, 1\ 800, 1\ 924, 1\ 1108\};$

$\boldsymbol{y} = \{18, 19, 24, 19, 31, 25, 50, 64, 68, 71\};$

$\boldsymbol{X1} = \boldsymbol{X}[, 2];$

$\boldsymbol{XPI} = INV(\grave{\boldsymbol{X}} * \boldsymbol{X});$

$\boldsymbol{BETA} = \boldsymbol{XPI} * \grave{\boldsymbol{X}} * \boldsymbol{y};$

$\boldsymbol{H} = \boldsymbol{X} * \boldsymbol{XPI} * \boldsymbol{X}';$

$\boldsymbol{RESID} = (\boldsymbol{I}(10) - \boldsymbol{H}) * \boldsymbol{y};$

$\boldsymbol{YHAT} = \boldsymbol{X} * \boldsymbol{BETA};$

$\boldsymbol{XYHAT} = \boldsymbol{X}1||\boldsymbol{YHAT};$

*CALL PGRAF(**XYHAT**,'*','x','Predicted','Plot of Predicted Values');*

ODS PDF CLOSE;

Example 15.6 *Creating a matrix from a SAS data set and vice versa.*

Consider the following data set consisting of the age and height of each of nine individuals:

Age (years)	Height (inches)
25	68
34	70
29	69
44	62
56	71
62	65
45	71
19	66
55	72

The following SAS statements can be used to create a matrix from this data set:

```
DATA ONE;
INPUT Age Height @@;
CARDS;
25 68 34 70 29 69 44 62 56 71 62 65 45 71 19 66 55 72
RUN;
PROC IML;
USE ONE;
VARNAMES = {Age Height};
READ ALL INTO X [COLNAME = VARNAMES];
PRINT X;
```

Note that the first five lines are SAS statements, but do not belong to IML and should therefore end with the RUN; statement. They are needed to create a SAS data set. This process gives the following **X** *matrix:*

$$\boldsymbol{X} = \begin{bmatrix} Age & Height \\ \hline 25 & 68 \\ 34 & 70 \\ 29 & 69 \\ 44 & 62 \\ 56 & 71 \\ 62 & 65 \\ 45 & 71 \\ 19 & 66 \\ 55 & 72 \end{bmatrix}.$$

Let us now reverse the process by retrieving the SAS data set from the columns of the matrix **X***. The following statements are needed:*

```
PROC IML;
X={25 68, 34 70, 29 69, 44 62, 56 71, 62 65, 45 71, 19 66, 55 72};
VARNAMES = {Age Height};
CREATE ONE FROM X [COLNAME = VARNAMES];
APPEND FROM X;
PROC PRINT DATA = ONE;
RUN;
```

Note that the third statement could have been deleted and the fourth one could have been written as

```
CREATE ONE FROM X [COLNAME={AGE HEIGHT}];
```

Example 15.7 *Consider the spectral decomposition of the matrix* **A** *in Example 6.13.*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

The SAS code to do this decomposition is

```
PROC IML;
A={1 2 2,2 1 2,2 2 1};
CALL EIGEN(M,P,A);
PRINT M;
PRINT P;
```

From the SAS output we get the column vector **M** *of eigenvalues and the matrix* **P** *of corresponding orthonormal eigenvectors of* **A**:

$$\boldsymbol{M} = \begin{bmatrix} 5 \\ -1 \\ -1 \end{bmatrix},$$

$$\boldsymbol{P} = \begin{bmatrix} 0.5773503 & -0.588553 & -0.565926 \\ 0.5773503 & 0.7843825 & -0.226739 \\ 0.5773503 & -0.195829 & 0.7926649 \end{bmatrix}.$$

It can be verified that $\boldsymbol{A} = \boldsymbol{P} * diag(\boldsymbol{M}) * \grave{\boldsymbol{P}}$.

Example 15.8 *Consider the singular-value decomposition of the matrix* **A** *in Example 7.3.*

$$\boldsymbol{A} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

The SAS code needed for this decomposition is

```
PROC IML;
A = {2 0 1 1,0 2 1 1,1 1 1 1};
CALL SVD(P,D,R1,A);
PRINT P D R1;
```

Example 15.9 *Consider the balanced random two-way crossed classification model*

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

$i = 1, 2, 3$; $j = 1, 2, 3, 4$; $k = 1, 2, 3$, *where* $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$; *and* $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$; *all random effects are independent. In vector form, this model can be written using direct products as was done in Section 13.1.*

$$\boldsymbol{y} = (\boldsymbol{1}_3 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3)\mu + (\boldsymbol{I}_3 \otimes \boldsymbol{1}_4 \otimes \boldsymbol{1}_3)\boldsymbol{\alpha} + (\boldsymbol{1}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{1}_3)\boldsymbol{\beta} + (\boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{1}_3)(\boldsymbol{\alpha\beta}) + \boldsymbol{I}_{36}\epsilon.$$

The variance-covariance matrix, $\boldsymbol{\Sigma}$, *of* $\boldsymbol{y}$ *is of the form*

$$\boldsymbol{\Sigma} = \sigma_\alpha^2(\boldsymbol{I}_3 \otimes \boldsymbol{J}_4 \otimes \boldsymbol{J}_3) + \sigma_\beta^2(\boldsymbol{J}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{J}_3) + \sigma_{\alpha\beta}^2(\boldsymbol{I}_3 \otimes \boldsymbol{I}_4 \otimes \boldsymbol{J}_3) + \sigma_\epsilon^2\boldsymbol{I}_{36},$$

where $\boldsymbol{J}_n$ *is a matrix of ones of order* $n \times n$. *Let us now generate a random vector* $\boldsymbol{y}$ *of order* 36×1 *from* $N(\mu\boldsymbol{1}_{36}, \boldsymbol{\Sigma})$, *given that* $\mu = 20.5$, $\sigma_\alpha^2 = 30$, $\sigma_\beta^2 = 20$, $\sigma_{\alpha\beta}^2 = 10$, *and* $\sigma_\epsilon^2 = 1$. *To do so, we can generate a random vector* $\boldsymbol{z}$ *of size* 36×1 *from the standard normal distribution. The desired random vector* $\boldsymbol{y}$ *can then be generated using the relation* $\boldsymbol{y} = 20.5\boldsymbol{1}_{36} + \grave{\boldsymbol{U}}\boldsymbol{z}$, *where* $\boldsymbol{U}$ *is an upper triangular matrix such that* $\boldsymbol{\Sigma} = \grave{\boldsymbol{U}}\boldsymbol{U}$, *which can be obtained from using the ROOT function that performs the Cholesky decomposition of* $\boldsymbol{\Sigma}$. *The needed SAS statements to generate* $\boldsymbol{y}$ *are*

```
PROC IML;
M1=I(3)@J(4,4,1)@J(3,3,1);
M2=J(3,3,1)@I(4)@J(3,3,1);
M3=I(12)@J(3,3,1);
M4=I(36);
SIGMA=30*M1 + 20*M2 + 10*M3 + M4;
U=ROOT(SIGMA);
z=J(36,1,0);
START A;
DO I=1 TO 36;
zz=NORMAL(0);
z[i,1]=zz;
END;
FINISH;
RUN A;
y = 20.5 * J(36, 1, 1) + U` * z;
PRINT y;
```

Let us now consider the $\boldsymbol{P}_i$ *matrices associated with the sums of squares for the main effects* α_i, β_j, *the interaction effect* $(\alpha\beta)_{ij}$, *and the error term. These sums of squares are*

$\boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y}$, *where* $\boldsymbol{P}_1 = \boldsymbol{M}1/12 - \boldsymbol{J}(36, 36, 1)/36$
$\boldsymbol{y}'\boldsymbol{P}_2\boldsymbol{y}$, *where* $\boldsymbol{P}_2 = \boldsymbol{M}2/9 - \boldsymbol{J}(36, 36, 1)/36$
$\boldsymbol{y}'\boldsymbol{P}_3\boldsymbol{y}$, *where* $\boldsymbol{P}_3 = \boldsymbol{M}3/3 - \boldsymbol{M}1/12 - \boldsymbol{M}2/9 + \boldsymbol{J}(36, 36, 1)/36$
$\boldsymbol{y}'\boldsymbol{P}_4\boldsymbol{y}$, *where* $\boldsymbol{P}_4 = \boldsymbol{I}(36) - \boldsymbol{M}3/3$.

It is of interest to find the eigenvalues of $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$, $\boldsymbol{P}_4$. *The SAS statements needed for this purpose, which should be added to the above SAS code, are*

```
P1=M1/12 - J(36,36,1)/36;
P2=M2/9 - J(36,36,1)/36;
P3=M3/3 - M1/12 - M2/9 + J(36,36,1)/36;
P4=I(36) - M3/3;
CALL EIGEN(MM1,E1,P1);
PRINT MM1;
CALL EIGEN(MM2,E2,P2);
PRINT MM2;
CALL EIGEN(MM3,E3,P3);
PRINT MM3;
CALL EIGEN(MM4,E4,P4);
PRINT MM4;
```

The SAS output shows that $\boldsymbol{P}_1$ *has two nonzero eigenvalues equal to one,* $\boldsymbol{P}_2$ *has three nonzero eigenvalues equal to one,* $\boldsymbol{P}_3$ *has six nonzero eigenvalues equal to one, and* $\boldsymbol{P}_4$ *has 24 nonzero eigenvalues equal to one. This is in agreement with the fact that* $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$, $\boldsymbol{P}_4$ *are idempotent matrices whose ranks are the same as the numbers of degrees of freedom for the main effects, the interaction effect, and the error term, which are equal to 2, 3, 6, and 24, respectively.*

Example 15.10 *An experiment was conducted to study the effects of temperature and curing time on the shear strength of the bonding of galvanized steel bars with a certain adhesive. The design used is a* 3×3 *factorial with three levels of temperature and three levels of time. Let* x_1 *and* x_2 *denote the coded levels of temperature and time, respectively. The experiment was carried out over a period of four dates where on a given date, batches of steel aliquots were selected at random from the warehouse supply. The same basic* 3×3 *design was used on each of the four dates, except that on dates 1 and 4, four replications of the center point were taken. The data given in Table 15.1 represent a portion of the data set reported in Khuri (2006).*

Note that blocking was done by dates, thus the four dates represent four blocks with the block effect considered as random. Let n_i denote the size of the ith block ($i = 1, 2, 3, 4$). The model for this experiment is

$$y_u = \beta_0 + \boldsymbol{f}'(\boldsymbol{x}_u)\boldsymbol{\beta} + \boldsymbol{z}_u'\boldsymbol{\gamma} + \boldsymbol{f}'(\boldsymbol{x}_u)\boldsymbol{\Lambda}\boldsymbol{z}_u + \epsilon_u, \quad u = 1, 2, \ldots, n,$$

where $n = \sum_{i=1}^{4} n_i$ is the total number of observations, $\boldsymbol{x}_u$ is the setting of $\boldsymbol{x} = (\boldsymbol{x}_{u1}, \boldsymbol{x}_{u2})'$ at the uth experimental run, $z_u = (z_{u1}, z_{u2}, z_{u3}, z_{u4})'$, where z_{ui} is an indicator variable taking the value 1 if the uth trial is in the ith block and 0 otherwise ($i = 1, 2, 3, 4; u = 1, 2, \ldots, n$), and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)'$ with γ_i denoting the effect of the ith block. The matrix $\boldsymbol{\Lambda}$ contains interaction coefficients between the blocks and the fixed polynomial terms in the model,

TABLE 15.1 Design Settings and Shear Strength Data (in psi)

x_1	x_2	Date One	Date Two	Date Three	Date Four
−1	−1	979	1151	1261	1288
0	−1	1777	1652	1859	1999
2	−1	2319	1897	2129	2264
−1	0	1586	1722	1548	1568
0	0	2371	2267	2142	2365
0	0	2351	2267	2142	2287
0	0	2182	2267	2142	2410
0	0	2181	2267	2142	2224
2	0	2551	2156	2402	2484
−1	1	1751	1511	1819	1712
0	1	2621	2389	2503	2548
2	1	2568	2465	2705	2812

and ϵ_u is a random experimental error. Furthermore, the mean of y_u is assumed to be a second-degree polynomial in x_1 and x_2 of the form

$$\beta_0 + \boldsymbol{f}'(\boldsymbol{x}_u)\boldsymbol{\beta} = \beta_0 + \beta_1 x_{u1} + \beta_2 x_{u2} + \beta_{11} x_{u1}^2 + \beta_{22} x_{u2}^2 + \beta_{12} x_{u1} x_{u2},$$

where $\beta_0, \beta_1, \beta_2, \beta_{11}, \beta_{22}$, and β_{12} are fixed unknown parameters. This model can be written in vector form as

$$\boldsymbol{y} = \beta_0 \mathbf{1}_n + \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \sum_{j=1}^{5} \boldsymbol{U}_j \boldsymbol{\delta_j} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is the vector of response values, $\mathbf{1}_n$ is a vector of n ones, $\boldsymbol{X}$ is a matrix of order $n \times 5$ and rank 5 whose uth row is $\boldsymbol{f}'(\boldsymbol{x}_u)$ $(u = 1, 2, \ldots, n)$, $\boldsymbol{Z} = \text{diag}(\mathbf{1}_{n_1}, \mathbf{1}_{n_2}, \mathbf{1}_{n_3}, \mathbf{1}_{n_4})$, $\boldsymbol{U}_j$ is a matrix of order $n \times 4$ whose ith column is obtained by multiplying the elements of the jth column of $\boldsymbol{X}$ with the corresponding elements of the ith column of $\boldsymbol{Z}$ $(i = 1, 2, 3, 4; j = 1, 2, 3, 4, 5)$, $\boldsymbol{\delta_j}$ is a vector of interaction coefficients between the blocks and the jth polynomial term $(j = 1, 2, 3, 4, 5)$, and $\boldsymbol{\epsilon}$ is the vector of the ϵ_u's $(u = 1, 2, \ldots, n)$.

The SAS statements needed to obtain the $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3, \boldsymbol{U}_4, \boldsymbol{U}_5$ matrices as well as the residual sum of squares are

```
DATA ONE;
INPUT BLOCK x1 x2 y @@;
CARDS;
1 -1 -1 979 1 0 -1 1777 1 2 -1 2319 1 -1 0 1586 1 0 0 2371 1 0 0 2351 1 0 0 2182 1 0 0
2181
1 2 0 2551 1 -1 1 1751 1 0 1 2621 1 2 1 2568
2 -1 -1 1151 2 0 -1 1652 2 2 -1 1897 2 -1 0 1722 2 0 0 2267 2 2 0 2156 2 -1 1 1511 2 0
1 2389 2 2 1 2465
3 -1 -1 1261 3 0 -1 1859 3 2 -1 2129 3 -1 0 1548 3 0 0 2142 3 2 0 2402 3 -1 1 1819 3 0
1 2503 3 2 1 2705
4 -1 -1 1288 4 0 -1 1999 4 2 -1 2264 4 -1 0 1568 4 0 0 2365 4 0 0 2287 4 0 0 2410 4 0 0
2224
4 2 0 2484 4 -1 1 1712 4 0 1 2548 4 2 1 2812
```

```
DATA TWO;
SET ONE;
x11 = x1 * x1;
x22 = x2 * x2;
x12 = x1 * x2;
RUN;
PROC IML;
USE TWO;
READ ALL INTO XX;
y = XX[1 : 42, 4];
PRINT y;
X = XX[1 : 42, {2 3 5 6 7}];
XXX = J(42, 1, 1)||X;
ONE1 = J(12, 1, 0);
ONE2 = J(9, 1, 1);
Z = BLOCK(ONE1, ONE2, ONE2, ONE1);
U1 = Z#X[1 : 42, 1];
U2 = Z#X[1 : 42, 2];
U3 = Z#X[1 : 42, 3];
U4 = Z#X[1 : 42, 4];
U5 = Z#X[1 : 42, 5];
G = XXX||Z||U1||U2||U3||U4||U5;
SSE = y` * [I(42) − G * GINV((G` * G)) * G`] * y;
PRINT X;
PRINT SSE;
```

Example 15.11 *Consider the random one-way model*

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, r; j = 1, 2, \ldots, n_i$$

where α_i and ϵ_{ij} are independently distributed as normal variates with zero means and variances, σ^2_α and σ^2_ϵ, respectively. Let $\bar{y}_{i.}$ and $\bar{\epsilon}_{i.}$ be the averages of the y_{ij}'s and ϵ_{ij}'s over j ($j = 1, 2, \ldots, n_i$). We then have

$$\bar{\boldsymbol{y}} = \mu \mathbf{1}_r + \boldsymbol{\alpha} + \bar{\boldsymbol{\epsilon}},$$

where $\bar{\boldsymbol{y}}$, $\boldsymbol{\alpha}$, and $\bar{\boldsymbol{\epsilon}}$ are vectors whose elements are $\bar{y}_{i.}$, α_i, and $\bar{\epsilon}_{i.}$, respectively, and $\mathbf{1}_r$ is the vector of ones of order $r \times 1$. The variance-covariance matrix of $\bar{\boldsymbol{y}}$ is

$$var(\bar{\boldsymbol{y}}) = \sigma^2_\alpha \boldsymbol{I}_r + \sigma^2_\epsilon \boldsymbol{K},$$

where $\boldsymbol{I}_r$ is the identity matrix of order $r \times r$, and

$$\boldsymbol{K} = diag\left(\frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_r}\right).$$

Consider the linear transformation, $\boldsymbol{u} = \boldsymbol{P}_1\bar{\boldsymbol{y}}$, *where the rows of* $\boldsymbol{P}_1$ *are orthonormal such that*

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{r}}\boldsymbol{1}'_r \\ \boldsymbol{P}_1 \end{bmatrix}$$

is an orthogonal matrix of order $r \times r$. *Then,* $\boldsymbol{u}$ *is normally distributed with a mean* $\boldsymbol{0}$ *and a variance-covariance matrix given by*

$$var(\boldsymbol{u}) = \sigma^2_\alpha \boldsymbol{I}_{r-1} + \sigma^2_\epsilon \boldsymbol{L},$$

where $\boldsymbol{L} = \boldsymbol{P}_1\boldsymbol{K}\boldsymbol{P}'_1$. *Note that the rows of* $\boldsymbol{P}_1$ *are eigenvectors of the matrix* $\boldsymbol{I}_r - \frac{1}{r}\boldsymbol{J}_r$ *with eigenvalue 1 of multiplicity* $r-1$. *The unweighted sum of squares for* α_i *is* $n_h \sum_{i=1}^r (\bar{y}_{i.} - \bar{y}^*)^2$, *where* $\bar{y}^* = (\frac{1}{r}) \sum_{i=1}^r \bar{y}_{i.}$ *and* n_h *is the harmonic mean of the* n_i*'s given by*

$$n_h = \frac{r}{\sum_{i=1}^r \frac{1}{n_i}}.$$

It can be shown that [*see Khuri(2002), Section 2*]

$$n_h \sum_{i=1}^r (\bar{y}_{i.} - \bar{y}^*)^2 = n_h \boldsymbol{u}'\boldsymbol{u}.$$

Let us consider the special case where $r = 6$ *and the design used is D with the following* n_i *values:* $D = \{3, 4, 4, 7, 9, 9\}$. *Its measure of imbalance according to Khuri (1996) is*

$$\phi = \frac{(\sum_{i=1}^6 n_i)^2}{r\sum_{i=1}^r n_i^2}.$$

The SAS statements needed to incorporate some of the above information are

```
PROC IML;
D = {3 4 4 7 9 9};
K1 = DIAG(D);
K = INV(K1);
ONE = J(1,6,1);
NH = 6/(ONE*K*ONE‘);
PHI = ((D*j(6,1,1))**2)/(6*(j(1,6,1)*(K1**2)*j(6,1,1)));
PRINT PHI;
MAT1 = I(6) - (1/6)*J(6,6.1);
CALL EIGEN(M1,E1,MAT1);
PRINT M1;
PRINT E1;
PP = E1[1:6, {1 2 3 4 5}];
P1 = PP‘;
PRINT P1;
```

```
L = P1*K*P̀1;
MAT2 = (I(6) - (1/6)*J(6,6,1))*K;
CALL EIGEN(M2,E2,MAT2);
PRINT M2;
CALL EIGEN(M3,E3,L);
PRINT M3;
```

Note that the nonzero eigenvalues of $\mathbf{L}$ *are the same as the nonzero eigenvalues of* $\mathbf{P}_1'\mathbf{P}_1\mathbf{K} = (\mathbf{I}_6 - (1/6) * \mathbf{J}(6,6,1)) * \mathbf{K} = MAT2$ *(see Theorem 6.12). Note also that the measure of imbalance* ϕ *is such that* $1/r < \phi \leq 1$. *A small value of* ϕ *indicates a high degree of imbalance. A data set is completely balanced when* $\phi = 1$. *In our special case where* $D = \{3,4,4,7,9,9\}$, $\phi = 0.8571$, *hence the data set has a small degree of imbalance. The eigenvalues of* $\mathbf{L}$ *are useful in assessing the adequacy of the method of unweighted means in providing approximate F-tests for an unbalanced random model with imbalance occurring only at the last stage[see Khuri (2002)].*

Example 15.12 *The effects of nitrogen and phosphorus fertilizers on irrigated corn yield were studied. The fertilizers were applied to experimental plots and the yield response was measured in pounds per plot. The level of the amount (lb/plot) of each fertilizer applied to a plot was determined by the coded settings of a central composite design. The design settings and harvested yields are shown in the following table:*

x_1 (Nitrogen)	x_2 (Phosphorus)	Yield
−1	−1	115
1	−1	170
−1	1	165
1	1	190
−1.414	0	118
1.414	0	180
0	−1.414	110
0	1.414	195
0	0	190

A second-degree model in x_1 *and* x_2 *was fitted to the yield data. The resulting prediction equation is*

$$\hat{y} = 189.994 + 20.962x_1 + 23.777x_2 - 7.5x_1x_2 - 18.187x_1^2 - 16.437x_2^2,$$

where $\hat{y}$ *is the predicted yield. It was also decided to do the following plots: the residuals versus the predicted yield values, the residuals versus* x_1, *the residuals versus* x_2, *the yield versus the predicted values, the yield versus* x_1, *and the predicted yield versus* x_1. *The SAS statements needed are*

```
DATA CORN;
INPUT X1 X2 YIELD @@;
X1SQ = X1*X1;
X2SQ = X2*X2;
X1X2 =X1*X2;
```

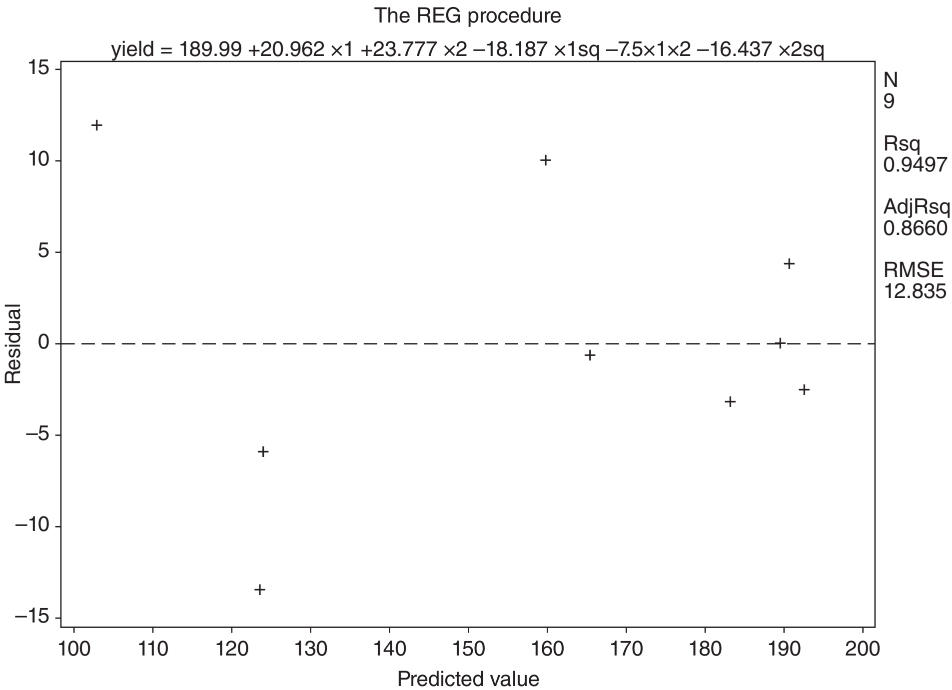


Figure 15.3 *Plot of Residual Versus Predicted Values.*

```
CARDS;
-1 -1 115 1 -1 170 -1 1 165 1 1 190 -1.414 0 118 1.414 0 180 0 -1.414 110 0 1.414 195
0 0 180
PROC PRINT DATA=CORN;
PROC REG DATA=CORN;
MODEL YIELD =X1 X2 X1SQ X1X2 X2SQ/p;
PLOT RESIDUAL.*PREDICTED.;
PLOT RESIDUAL.*X1;
PLOT RESIDUAL.*X2;
PLOT YIELD*PREDICTED.;
PLOT YIELD*X1;
PLOT PREDICTED.*X1;
RUN;
```

Some of the plots are displayed in Figures 15.3, 15.4, 15.5, and 15.6.

These plots provide diagnostic tools for checking the adequacy of the fitted model. In Figure 15.3, the residuals are plotted against the predicted values. If the plot shows a random scatter of points contained within a horizontal band, then the model is considered adequate. In our case, the points appear to be mostly at random with the possibility of one or two outliers. Plots of the residuals against the control variables x_1 *and* x_2 *are useful in determining if the second-degree model is an adequate representative of the yield response. This can be the case if the residuals fall within a horizontal band. If, however, the residuals exhibit a particular pattern, then the addition of higher-order terms in* x_1 *and* x_2 *to the model may be needed. The plot in Figure 15.4 does not appear to have any particular*

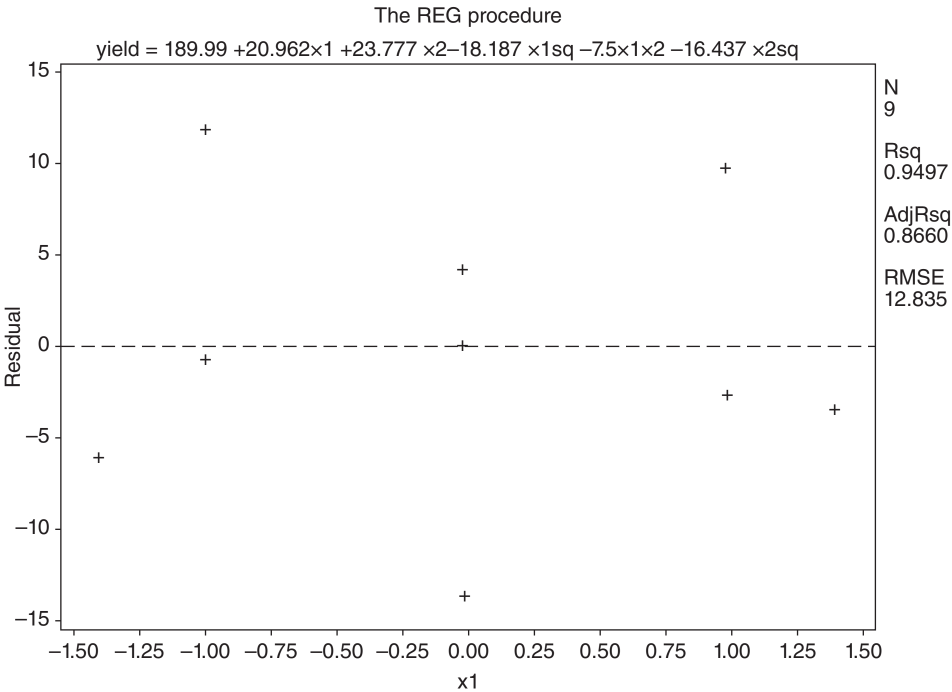


Figure 15.4 *Plot of Residuals Versus* x_1.

pattern. However, the plot in Figure 15.5 exhibits some upward curvature with the possibility of one outlier. The plot of the observed yield against its predicted value as in Figure 15.6 is helpful in determining if there is an agreement between the two. This can be the case if the points line up on a straight line through the origin at an angle of 45° *with the horizontal axis. If, however, the plot appears to have some curvature, then a higher-degree model may be needed. In our case, the straight line through the origin appears to exhibit a reasonable fit to the data points with the possibility of one or two outliers.*

Example 15.13 *Let us again consider the corn data used in Example 15.12. It is desirable to measure the influence of each observation in this data set. This can be accomplished by using Cook's distance measure; Cook (1977, 1979) used a measure of the squared distance between the least-squares estimate based on all the observations and the estimate obtained by deleting the ith observation. It is known [see, for example, Montgomery et al. (2012, Section 6.3)] that the expression for* D_i *for the ith observation is given by*

$$D_i = \frac{(\hat{\boldsymbol{y}}_{(i)} - \hat{\boldsymbol{y}})'(\hat{\boldsymbol{y}}_{(i)} - \hat{\boldsymbol{y}})}{pMS_{Res}},$$

where $\hat{\boldsymbol{y}}_{(i)}$ *is the predicted response vector obtained after deleting the ith observation,* $i = 1, 2, \ldots, n$*, where n is the number of observations, p is the number of parameters in the model, and* MS_{Res} *is the residual mean square.*

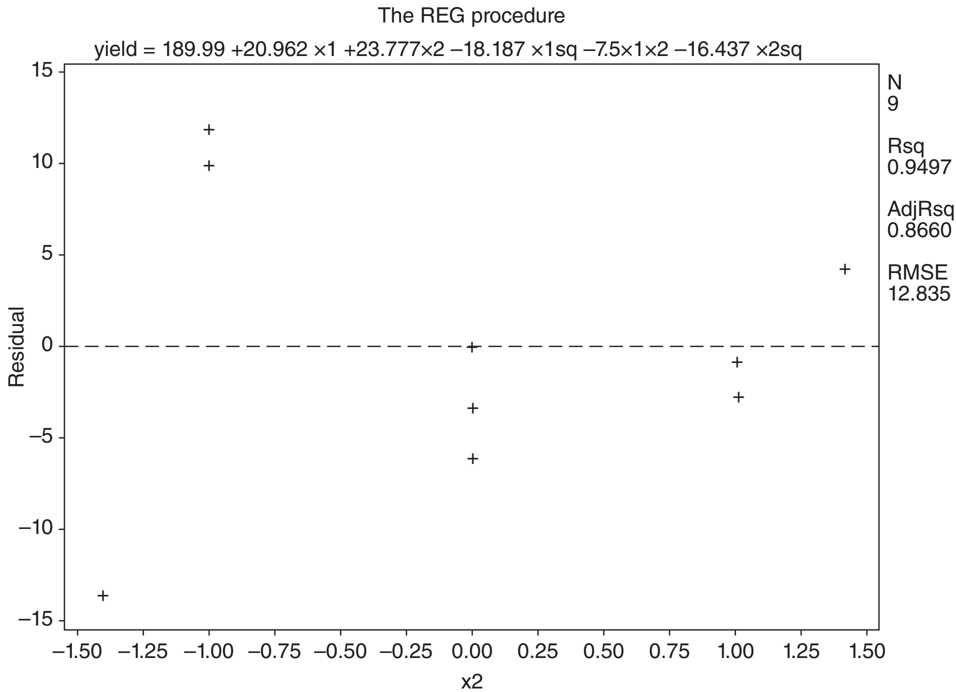


Figure 15.5 *Plot of Residuals Versus x_2.*

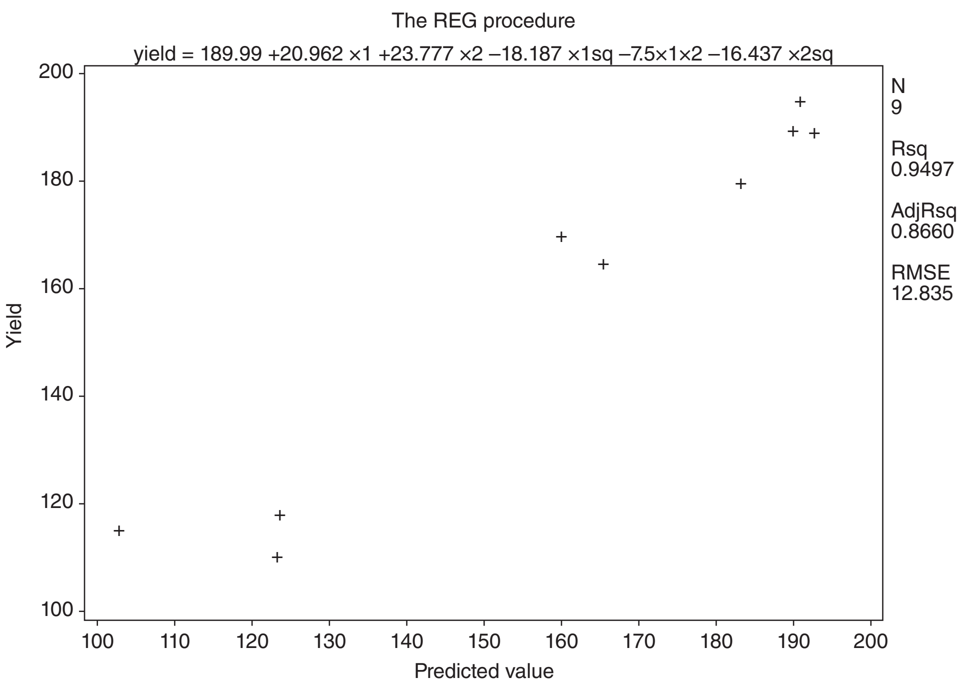


Figure 15.6 *Plot of Yield Versus Predicted Values.*

The SAS statements needed to get values of D_i are
PROC REG DATA=CORN;
MODEL YIELD = X1 X2 X1SQ X1X2 X2SQ/r;
TITLE 'Regression for Quadratic Model:Corn Data';
RUN;

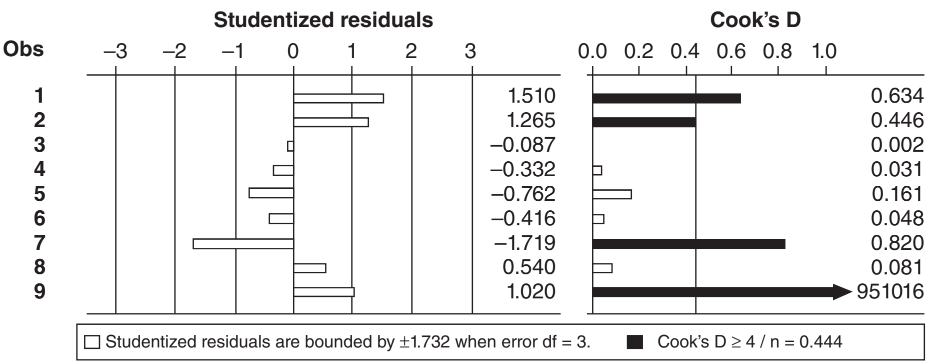
where the *r* option in the model statement specifies the residual. The output produced as a result of requesting the *r* option is shown in Figure 15.7. Observations for which $D_i > 1$

Regression for quadratic model: corn data

The REG procedure
model: MODEL1
dependent variable: yield

Output statistics							
Obs	**Dependent variable**	**Predicted value**	**Std error mean predict**	**Residual**	**Std error residual**	**Student residual**	**Cook's D**
1	115	103.1315	10.1474	11.8685	7.859	1.510	0.634
2	170	160.0548	10.1474	9.9452	7.859	1.265	0.445
3	165	165.6861	10.1474	–0.6861	7.859	–0.087	0.002
4	190	192.6095	10.1474	–2.6095	7.859	–0.332	0.031
5	118	123.9911	10.1462	–5.9911	7.860	–0.762	0.161
6	180	183.2707	10.1462	–3.2707	7.860	–0.416	0.048
7	110	123.5098	10.1462	–13.5098	7.860	–1.719	0.820
8	195	190.7521	10.1462	4.2479	7.860	0.540	0.081
9	190	189.9944	12.8348	0.005594	0.00548	1.020	951015.8

Studentized residuals and cook's D for yield



Sum of residuals	0
Sum of squared residuals	494.19929
Predicted residual SS (Press)	939986342

Figure 15.7 *Studentized Residuals and Cook's D for Yield.*

are considered influential, that is, their removal can affect predictions and the estimated parameter vector. We note that observation 9 has a very large D_9 which indicates that it is highly influential.

The following references are useful in getting more information about the use of SAS: Cody (2011), Cody and Smith (2006), Littell et al. (2002), Perrett (2010), Schlotzhauer and Littell (1997).

15.5 EXPLANATIONS OF SAS STATEMENTS USED EARLIER IN THE TEXT

1. The SAS statements to get the least-squares estimates of the parameters in model (11.36) are
```
DATA;
INPUT X1 X2 X3 Y;
CARDS;
(include here the data set in Table 11.3)
PROC GLM;
MODEL Y=X1 X2 X3 X1*X2 X1*X3 X2*X3 X1*X1 X2*X2 X3*X3/p XPX CLM;
RUN;
```
The p option in the model statement prints observed, predicted, and residual values for each observation in the data set. The XPX option prints the $\boldsymbol{X}'\boldsymbol{X}$ matrix. The CLM option prints confidence limits for a mean predicted value for each observation.
Note: The GLM in PROC GLM stands for General Linear Models. It is used for many different analyses including regression, analysis of variance, analysis of covariance, response surface models, multivariate analysis of variance, and repeated measures analysis of variance, among others.
2. The SAS statements used in Example 11.6: The first six statements are used to get the least-squares estimates of the parameters in the simple linear regression model. The OUTPUT OUT statement creates a new data set called “NEW” which contains the predicted weight values. Statements 9–10 are used to plot observed weight values against height and predicted weight values against height, and the two plots are superimposed on top of each other on one set of axes.
3. The SAS statements used in Example 11.7: These statements give the least-squares estimates of the parameters in model (11.36). Furthermore, the first PROC GLM statement is used to get the confidence intervals on the mean response for each observation which results from including the CLM option in the MODEL statement. The second PROC GLM is used to get the prediction intervals, that is, confidence limits for individual predicted values for each observation. Note that CLI and CLM should not be used together as options in the same MODEL statement.
4. The SAS statements used in Exercise 4 of Chapter 11: PROC GLM is used to get the least-squares estimates of the model parameters and to create Data set Two which contains Data set One. The USE statement opens Data set Two for reading, and the READ ALL INTO statement reads all observations from Data set Two turning them into columns of matrix X11. The matrix XX1 consists of columns 1 and 2 of X11, and the vector Y consists of column 3 of X11; MSE is the error mean square. BETA is the vector of parameter estimates, and F is the F-statistic value for testing the null hypothesis in part (b). LAMBDA is the noncentrality parameter. The POWER statement gives the power of the F-test for part (b). Note that the PROBF function gives

the probability that an observation from the F-distribution with 2 and 9 degrees of freedom is less than or equal to $x = 4.24$, the test statistic value.

5. The SAS statements for Exercise 6, Chapter 12: The E option in PROC GLM's model statement prints the general form of all estimable function. The SOLUTION option prints a solution to the normal equations (parameter estimates). Under PROC IML, MS_E, the error mean square and BETA, a vector of parameter estimates , are printed. The F statement gives the value of the test statistic for testing H_0 in part (b). The power of the test in part (c) is given by the POWER statement.
6. The SAS statements for Exercise 10, Chapter 12: The CLASS statement names the classification variables to be used in the analysis. Such a statement is used with any analysis of variance model. The E option in the model statement gives the results needed for parts (a), (b), and (c). The SOLUTION option gives a solution to the normal equation (parameter estimated).
7. The SAS statements used in Exercise 2, Chapter 13: The RANDOM statement specifies which effects in the model are random. The TEST option provides tests for each effect in the model.
8. The SAS statements used in Exercise 3, Chapter 14: CALL EIGEN(**M,E,H**); computes the eigenvalues of $\boldsymbol{H} = \boldsymbol{G}\grave{\boldsymbol{G}}$, given by the elements of the column vector $\boldsymbol{M}$, and the matrix $\boldsymbol{E}$ whose columns are the corresponding orthonormal eigenvectors of $\boldsymbol{H}$ (recall that $\grave{\boldsymbol{G}}$ is the operator used in PROC IML to denote the transpose of $\boldsymbol{G}$).
9. The SAS statements used in Section 15.3: DATA TWO was created using DATA ONE. The former is used in PROC IML to create the columns of the matrix $\boldsymbol{G}$ which has three columns, namely, columns one and two give the levels of A and B and the third column gives the y values.
10. The SAS statements used in Example 15.1: The determinant, inverse, and trace of the matrix $\boldsymbol{A}$ are computed. In addition, the eigenvalues and eigenvectors of $\boldsymbol{A}$ are obtained using CALL EIGEN(**M,E,A**). The sum of all elements of $\boldsymbol{A}$ is given by value of s, and the sum of squares of all elements of $\boldsymbol{A}$ is given by the value of ss. The rank function is used to create the matrix $\boldsymbol{r}$ which contains the corresponding ranks of the matrix $\boldsymbol{A}$.
11. The SAS statements used in Example 15.2: CALL EIGEN(**M,E,A**) is used to obtain the eigenvalues and eigenvectors of the matrix $\boldsymbol{A}$. The ROOT(**A**) function gives the Cholesky decomposition of $\boldsymbol{A}$ such that $\boldsymbol{A} = \grave{\boldsymbol{U}}\boldsymbol{U}$, where $\boldsymbol{U}$ is an upper triangular matrix (see Theorem 6.3).
12. The SAS statements used in Example 15.3: The inverse of the matrix $\grave{\boldsymbol{X}}\boldsymbol{X}$, the least-squares estimate of $\boldsymbol{\beta}$, and the $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ matrix are computed. The residual vector RESID and the residual sum of squares and mean square are obtained, in addition to the estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, **VCBETA**.
13. The SAS statements used in Example 15.4: The inverse matrix of $\boldsymbol{X}'\boldsymbol{X}$, the vector of least-squares estimate of $\boldsymbol{\beta}$, **BETAHAT**, and its two elements, $BETAHAT_0$ and $BETAHAT_1$, are computed. The $\boldsymbol{H} = \boldsymbol{X}(\grave{\boldsymbol{X}}\boldsymbol{X})\grave{\boldsymbol{X}}$ matrix, the residual vector, RESID, and the error sum of squares and mean square are obtained. The vector of x values, XVALUE; the sum of squares of the x values, SQUAREX; the sum of the x values, SUMX; the average, XBAR of these values; and the quantity S_{xx} are computed. The lower and upper limits of the 95% confidence interval on β_1, as well as those on β_0 are computed. Furthermore, the lower and upper limits of the 95% confidence interval on the error variance, LCISIGMASQ and UCISIGMASQ, are obtained.

14. The SAS statements used in Example 15.5:
    (a) Statements used to generate Figure 15.1: The inverse of the matrix $\boldsymbol{X}'\boldsymbol{X}$, $\hat{\boldsymbol{\beta}}$, the matrix $\boldsymbol{H} = \boldsymbol{X}(\grave{\boldsymbol{X}}\boldsymbol{X})^{-1}\grave{\boldsymbol{X}}$, and the residual vector are obtained. The two-column matrix $\boldsymbol{XR}$ consists of the second column of the matrix $\boldsymbol{X}$ (values of x) and the residual vector. CALL PGRAF is used to produce scatter plots of the x values versus the residuals. The points are labelled using the "r" symbol, and the title of the figure is specified. Note that the SAS Output Delivery System (ODS) enables one to manage graphics that are created by a SAS procedure. It is used to send output from the IML procedure to PDF files.
    (b) Statements used to generate Figure 15.2: These are similar to the statements used in part (a), except that the x values are plotted against the predicted response values.
15. The SAS statements used in Example 15.6: DATA ONE, which consists of the values of Age and Height, are defined. The USE statement opens DATA ONE for reading. The variables are named according to the VARNAMES statement. The READ ALL INTO **X** [COLNAME=VARNAMES] statement reads all observations from DADA ONE into the matrix $\boldsymbol{X}$ whose columns are named according to the VARNAMES statement.
    The second group of SAS statements used in Example 15.6: The CREATE statement creates a new data set called ONE from the matrix $\boldsymbol{X}$. The names, Age and Height, are associated with the new data set. The APPEND statement adds observations from the matrix $\boldsymbol{X}$ containing the data to append. Note that the RUN statement was added at the end since it followed PROC PRINT which is not an IML statement.
16. The SAS statements used in Example 15.7: CALL EIGEN(**M**,**P**,**A**) computes the eigenvalues and eigenvectors of the matrix $\boldsymbol{A}$. The eigenvalues are real since $\boldsymbol{A}$ is symmetric. They are given by the elements of the column vector $\boldsymbol{M}$ in a descending order. The eigenvectors of $\boldsymbol{A}$ are given by the columns of the matrix $\boldsymbol{P}$ which is orthogonal. The columns are arranged so that the first one corresponds to the largest eigenvalue, and so forth.
17. The SAS statements used in Example 15.8: CALL SVD(**P**,**D**,$\boldsymbol{R_1}$,**A**) computes the singular-value decomposition of the matrix $\boldsymbol{A}$ such that $\boldsymbol{A} = \boldsymbol{P}Diag(\boldsymbol{D})\boldsymbol{R}_1'$. Note first, that if in the formula (7.22) that gives the expression for the singular-value decomposition, the matrix $\boldsymbol{Q}$ is partitioned as $\boldsymbol{Q} = [\boldsymbol{R}_1 : \boldsymbol{R}_2]$, where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are matrices of orders $n \times m$ and $n \times n - m$, respectively, then formula (7.22) can be written as

$$\boldsymbol{A} = \boldsymbol{P}Diag(\boldsymbol{D})\boldsymbol{R}_1',$$

    which is the form of the singular-value decomposition given in PROC IML. Recall that the diagonal elements of the diagonal matrix $\boldsymbol{D}$ are the non-negative square roots of the eigenvalues of $\boldsymbol{A}\grave{\boldsymbol{A}}$.
18. The SAS statements used in Example 15.9: The ROOT(**SIGMA**) function performs the Cholesky decomposition of **SIGMA** such that $\boldsymbol{SIGMA} = \grave{\boldsymbol{T}}\boldsymbol{T}$, where $\boldsymbol{T}$ is upper triangular (see Theorem 6.3). The SAS module defined by the statements START A through FINISH is used to generate 36 random variables having the standard normal distribution. The vector $\boldsymbol{y}$ is defined as shown in the $\boldsymbol{y}$ statement. The eigenvalues of the matrices $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ are computed using CALL EIGEN.
19. The SAS statements used in Example 15.10: DATA ONE is defined. DATA TWO is defined using DATA ONE. DATA TWO is used to read the variables into the columns

of the matrix $\boldsymbol{XX}$. The vector $\boldsymbol{y}$ is defined as the fourth column of $\boldsymbol{XX}$. The matrix $\boldsymbol{X}$ is defined and consists of columns 2, 3, 5, 6, 7 of $\boldsymbol{XX}$. The matrix $\boldsymbol{XXX}$ is obtained by horizontally joining $\mathbf{1}_{42}$ with $\boldsymbol{X}$. The matrix $\boldsymbol{Z}$ is defined. The $\boldsymbol{U}_i$ matrices are obtained by performing products of the corresponding elements of the matrix $\boldsymbol{Z}$ with each of the columns 1,2,3,4,5 of $\boldsymbol{X}$. The matrix $\boldsymbol{G}$ is then defined and the residual sum of squares SS_E is computed.

20. The SAS statements used in Example 15.11: The design **D** is defined as a row vector. The diagonal elements of the diagonal matrix $\boldsymbol{K}_1$ consists of the elements of **D**. The inverse matrix $\boldsymbol{K}$ of $\boldsymbol{K}_1$ is defined. The harmonic mean NH and measure of imbalance PHI are computed. The matrix **MAT1** is defined and its eigenvalues **M1** and eigenvectors **E1** are obtained using CALL EIGEN. The matrix ***PP*** consists of columns 1,2,3,4, and 5 of the matrix **E1**. The matrix **MAT2** is defined as $(\boldsymbol{I}_6 - \frac{1}{6}\boldsymbol{J}_6)\boldsymbol{K}$ whose eigenvalues are the same as those of the matrix $\boldsymbol{L}$.
21. The SAS statements used in Example 15.12: The data set named Corn is defined. Note that the model statement in PROC REG does not allow quadratic or higher-order terms. To include such terms, they must be defined separately after the input statement such as X1SQ, X2SQ, and X1X2 and then added to the model. The various plot statements are for residuals versus predicted values, residuals versus x_1, residuals versus x_2, yield versus predicted yield, yield versus x_1, and predicted values versus x_1.
22. The SAS statements used in Example 15.13: The use of the residual option, r, in the model statement results in the values of the studentized residuals (obtained by dividing the residuals by their standard errors) and their corresponding Cook's D_i's as shown in Figure 15.7. Observations for which $D_i > 1$ are considered to be influential.

REFERENCES

Cody, R. (2011). *SAS Statistics by Example*. SAS Institute Inc., Cary, NC.

Cody, R. and Smith, J. K. (2006). *Applied Statistics and the SAS Programming Language*, 5th ed. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19, 15–18.

Cook, R. D. (1979). Influential observations in linear regression. *J. Am. Stat. Assoc.*, 74, 169–174.

Khuri, A. I. (1996). A method for determining the effect of imbalance. *J. Stat. Plann. Inf.*, 55, 115–129.

Khuri, A. I. (2002). Graphical evaluation of the adequacy of the method of unweighted means. *J. Appl. Stat.*, 29, 1107–1119.

Khuri, A. I. (2006). Mixed response surface models with heterogeneous within-block error variances. *Technometrics*, 48, 206–218.

Littell, R. C., Stroup, W.W., and Freund, R.J. (2002). *SAS for Linear Models*, 4th ed. SAS Institute Inc., Cary, NC.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis*, 5th ed. John Wiley & Sons, Hoboken, NJ.

Perrett, J. J. (2010). *A SAS/IML Companion for Linear Models*. Springer, New York.

SAS Institute Inc. (2004). *SAS/IML 9.1 User's Guide*. SAS Institute Inc., Cary, NC.

Schlotzhauer, S. D. and Littell, R. C. (1997). *SAS System for Elementary Statistical Analysis*, 2nd ed. SAS Institute Inc., Cary, NC.

EXERCISES

15.1 Consider the matrix $\mathbf{A}$ given by

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

(a) Demonstrate that $\mathbf{A}$ is nonsingular.

(b) Find the eigenvalues of $\mathbf{A}$ and demonstrate that it is positive definite.

(c) Find the rank of $\mathbf{A}$ and demonstrate that this rank is equal to the rank of $\mathbf{A}^2$.

15.2 Consider again the matrix $\mathbf{A}$ in Exercise 1. Let m and s be scalars defined as

$$m = \frac{tr(\mathbf{A})}{3}, \qquad s = \left[\frac{tr(\mathbf{A}^2)}{3} - m^2\right]^{1/2}.$$

(a) Find the inverse of $\mathbf{A}$ and the trace of $\mathbf{A}^2$.

(b) Show that $e_{max}(\mathbf{A}) \leq m + s\sqrt{2}$, where $e_{max}(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$.

(c) Show that $e_{max}(\mathbf{A}) - e_{min}(\mathbf{A}) \leq s\sqrt{6}$, where $e_{min}(\mathbf{A})$ is the smallest eigenvalue of $\mathbf{A}$.

15.3 Consider the first-degree model, which is fitted using the data shown below.

y	x_1	x_2
30	1.4	1.5
61	4.1	2.1
64	5.5	1.9
16.9	2.1	.80
95	7.8	3.2
76	6.5	2.9
93.3	12.8	1.5
64.4	7.7	1.7
93.9	12.1	1.2
81.3	6.9	2.2
77.5	6.1	2.9
65.9	8.2	1.4
77.8	7.9	2.3
74.6	8.5	2.1
61.7	7.7	1.3
84.1	11.6	1.2
42.2	6.5	.80
55.9	2.1	2.6
31.9	4.8	.84
66.1	7.1	1.2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

(a) Give a plot of the residuals against x_1 and x_2.

(b) Give a plot of the residuals versus the predicted values.

(c) What comments can be made based on the plots in (a) and (b)?

15.4 Consider again the y values in Exercise 3. It is desirable to determine if these values come from a normal distribution. For this reason, the *normal probability plot* in PROC UNIVARIATE can be used. The following statements are needed:

```
DATA ONE;
INPUT Y @@;
CARDS;
(include the y values here)
PROC UNIVARIATE DATA=ONE PLOT;
VAR Y;
TITLE 'NORMALITY PLOT';
RUN;
```

If the points of the plot lie approximately on a straight line, then this gives an indication of a normal distribution. Substantial departures from a straight line indicate that the distribution is not normal. On the basis of this information, determine if the y values can possibly come from a normal distribution.

15.5 Suppose that we have the following matrix:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 0 & 1 \\ 5 & 6 & 9 & 0 \\ 0 & 9 & 8 & 5 \\ 1 & 0 & 5 & 15 \end{bmatrix}$$

(a) Verify that $tr(\boldsymbol{A}) = \sum_{i=1}^{4} \lambda_i$, where λ_i is the ith eigenvalue of $\boldsymbol{A}$.

(b) Verify that $|\boldsymbol{A}| = \prod_{i=1}^{4} \lambda_i$.

15.6 Consider the data set in Exercise 3. Determine if there are any influential observations in this data set.

15.7 Consider the two matrices,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 1 \\ 0 & 1 & 17 \end{bmatrix}$$

(a) Show that $\boldsymbol{B}$ is positive definite.

(b) Find a lower bound and an upper bound on the ratio $\dfrac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}}$.

15.8 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 2 & 3 \\ 5 & 3 & 1 \end{bmatrix}$$

Let m and s be defined as:

$$m = \frac{tr(\boldsymbol{A})}{3}, \quad s = \left[\frac{tr(\boldsymbol{A}^2)}{3} - m^2\right]^{1/2}.$$

(a) Verify that

$$m - s\sqrt{n-1} \leq e_{min}(\boldsymbol{A}) \leq m - \frac{s}{\sqrt{n-1}},$$

where $e_{min}(\boldsymbol{A})$ denotes the smallest eigenvalue of $\boldsymbol{A}$.

(b) Verify that

$$m + \frac{s}{\sqrt{n-1}} \leq e_{max}(\boldsymbol{A}) \leq m + s\sqrt{n-1},$$

where $e_{max}(\boldsymbol{A})$ denotes the largest eigenvalue of $\boldsymbol{A}$.

(c) Verify that

$$e_{max}(\boldsymbol{A}) - e_{min}(\boldsymbol{A}) \leq s\sqrt{2n}.$$

15.9 Consider the following matrices:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 5 \\ 5 & 0 & 1 \\ 8 & 3 & 2 \\ 1 & 0 & 9 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 3 & 7 & -1 \\ -2 & 5 & 0 \\ 1 & 8 & 3 \\ 6 & -2 & 1 \end{bmatrix}.$$

Verify that

$$[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \leq tr(\boldsymbol{A}'\boldsymbol{A})\, tr(\boldsymbol{B}'\boldsymbol{B}),$$

which is the Cauchy–Schwarz inequality for matrices.

15.10 Let $\boldsymbol{A}$ be an $n \times n$ matrix.

(a) Show that $tr(\boldsymbol{A}^2) \leq tr(\boldsymbol{A}\boldsymbol{A}')$.

(b) Verify the inequality in (a) using the matrix

$$\boldsymbol{A} = \begin{bmatrix} 3 & 0 & -1 \\ 6 & 7 & 3 \\ 5 & 1 & 8 \end{bmatrix}.$$

15.11 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ symmetric matrices.

(a) Show that $tr[(\boldsymbol{A}\boldsymbol{B})^2)] \leq tr(\boldsymbol{A}^2\boldsymbol{B}^2)$.

(b) Verify the result in (a) using the matrices

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 7 & 3 \\ -1 & 3 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -3 & 5 \\ -3 & 8 & 6 \\ 5 & 6 & 1 \end{bmatrix}.$$

15.12 **(a)** Let A be a positive definite matrix of order $n \times n$. Show that

$$(|A|)^{1/n} \leq \frac{1}{n} tr(A).$$

(b) Let A be defined as

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & 1 \\ -1 & 1 & 12 \end{bmatrix}$$

Show that A is positive definite, then use it to verify the inequality in (a).

16

Use of MATLAB in Matrix Computations

MATLAB is a software for technical computing which also includes a graphic capability and a programming environment. MATLAB stands for MATrix LABoratory. It has its own language that makes it easier to execute math-related tasks. It also has graphics commands which facilitate the visualization of results. MATLAB runs on Windows, Macintosh, and Unix machines, to name just a few.

After installing MATLAB on the computer system, typing the command MATLAB at the MATLAB prompt results in activating it. Alternatively, one can activate MATLAB by double clicking on its icon. Exiting MATLAB can be done by typing the command "quit" or "exit". To save a file created in the workspace (the area of memory accessible from the command line), one can use the command "save filename", where "filename" is an arbitrary file name. To retrieve a file that has already been saved, one can enter the command "load filename".

Note that MATLAB is case sensitive and all variables must start with a letter. Comments can be added to MATLAB statements by using the symbol %, for example, % the variable x will be defined later.

16.1 ARITHMETIC OPERATORS

There are two types of arithmetic operators: array operators and matrix operators. The array operators are done element-by-element whereas matrix operators are the usual operators used in matrix algebra. The array operators are distinguished by using the period character

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

TABLE 16.1 Arithmetic Operators

Operator	Function of Operator
+	Sum of scalars, vectors, or matrices
−	Subtraction of scalars, vectors, or matrices
*	Product of scalars or matrices
. *	Array product. For two matrices of the same order, $\boldsymbol{A}$,$\boldsymbol{B}$, if $\boldsymbol{C} = \boldsymbol{A}. * \boldsymbol{B}$, then $\boldsymbol{C} = (a_{ij}b_{ij})$
/	Quotient of scalars or matrices by scalars. For two matrices, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{A}/\boldsymbol{B} = \boldsymbol{A} * inv(\boldsymbol{B})$, where $inv(\boldsymbol{B})$ is the inverse of $\boldsymbol{B}$
./	$\boldsymbol{A}./\boldsymbol{B} = (\frac{a_{ij}}{b_{ij}})$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of the same order
^	Power of a scalar or a matrix, for example, $\boldsymbol{A}^k$, where $\boldsymbol{A}$ is a matrix and k is a scalar
.^	Array power (element-wise power): $\boldsymbol{A}.^{\boldsymbol{B}} = (a_{ij}{}^{b_{ij}})$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of the same order

".". For example, If $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ are two matrices of the same order, then the command

$$\boldsymbol{C} = \boldsymbol{A}. * B$$

produces the matrix $\boldsymbol{C}$ of the same order with elements $c_{ij} = a_{ij} * b_{ij}$. Using the command

$$\boldsymbol{C} = \boldsymbol{A} * \boldsymbol{B}$$

results in the usual product of two matrices used in matrix algebra, that is, $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, where n is the number of columns of $\boldsymbol{A}$, which is equal to the number of rows of $\boldsymbol{B}$. A list of arithmetic operators is given in Table 16.1.

16.2 MATHEMATICAL FUNCTIONS

MATLAB has several predefined mathematical functions which can be evaluated by typing the name of a given function, for example, sin(x), exp(x), log(x). A list of some of these functions is shown in Table 16.2. In this table, note that in order to use sqrtm($\boldsymbol{A}$), $\boldsymbol{A}$ must be a square matrix and that $sqrtm(\boldsymbol{A})$ is a matrix such that $[sqrtm(\boldsymbol{A})] * [sqrtm(\boldsymbol{A})] = \boldsymbol{A}$. However, $sqrt(\boldsymbol{A})$ gives the square roots of the individual elements of $\boldsymbol{A}$ which does not have to be square. For example, for the matrix $\boldsymbol{A}$ given in (16.1), we have

$$sqrt(\boldsymbol{A}) = \begin{bmatrix} 2.2361 & 1.4142 & i & 2.6458 \\ 2.2361i & 0 & 3 & 2.6458 \\ 3 & 3.1623 & 2i & 3.1623 \end{bmatrix},$$

where i is the complex number $= \sqrt{(-1)}$.

TABLE 16.2 Mathematical Functions

Symbol	Function
sin(*x*)	Sine function
cos(*x*)	Cosine function
tan(*x*)	Tangent function
sec(*x*)	Secant function
csc(*x*)	Cosecant function
cot(*x*)	Cotangent function
log(*x*)	Natural logarithmic function (base *e*)
log10(*x*)	Common logarithmic function (to the base 10)
sqrt(*x*)	Square root function
abs(*x*)	Absolute value function
sqrtm(***X***)	Square root function for the square matrix ***X***. It is defined such that if $\boldsymbol{Y} = sqrtm(\boldsymbol{X})$, then $\boldsymbol{Y} * \boldsymbol{Y} = \boldsymbol{X}$.

16.3 CONSTRUCTION OF MATRICES

To construct an $m \times n$ matrix $\boldsymbol{A}$, we use square brackets at the beginning and end of the matrix. The elements of a row are separated by spaces. The rows of $\boldsymbol{A}$ are separated by semicolons. For example, consider the 3×4 matrix $\boldsymbol{A}$ written in MATLAB as

$$\boldsymbol{A} = [5 \quad 2 \quad -1 \quad 7\,; -5 \quad 0 \quad 9 \quad 7\,; 9 \quad 10 \quad -4 \quad 10],$$

which, in standard matrix notation, is written as

$$\boldsymbol{A} = \begin{bmatrix} 5 & 2 & -1 & 7 \\ -5 & 0 & 9 & 7 \\ 9 & 10 & -4 & 10 \end{bmatrix}. \tag{16.1}$$

Vectors can be constructed as matrices since they have only one row (for a row vector) or one column (for a column vector). There are two other ways to enter vectors for special purposes. For example, typing $\boldsymbol{u} = (0 : 0.10 : 2)$ gives a row vector that contains all the numbers between 0 and 2, inclusively, with an increment of 0.10. Another way to construct a vector is to use the command *linspace*. For example, typing $\boldsymbol{v} = linspace(-a1, a2)$ gives a row vector of 100 (default value) equally spaced points between $a1$ and $a2$. To use a different number of points, for example n points (different from 100), use $linspace(a1, a2, n)$ which produces a row vector of n points between $a1$ and $a2$, inclusively, that are equally spaced.

16.3.1 Submatrices

To view a particular element in a matrix, for example, the element in row 1 and column 2 in matrix $\boldsymbol{A}$, we type $\boldsymbol{A}(1, 2)$ which gives 2. To obtain a submatrix of a matrix, for example the submatrix consisting of rows 2, 3 and columns 2, 3 of matrix $\boldsymbol{A}$, we type $\boldsymbol{A}(2:3,2:3)$ which gives

$$\boldsymbol{A}(2:3,2:3)=\begin{bmatrix} 0 & 9 \\ 10 & -4 \end{bmatrix}.$$

To view a particular row of a matrix, for example, row 3 of $\boldsymbol{A}$, we type $\boldsymbol{A}(3,:)$ which gives $\boldsymbol{A}(3,:)=[9\ \ 10\ \ -4\ \ 10]$. Also, to view one column, or a group of columns, of a matrix, for example columns 2 and 3 of $\boldsymbol{A}$, we type $\boldsymbol{A}(:,2:3)$ which gives

$$\boldsymbol{A}(:,2:3)=\begin{bmatrix} 2 & -1 \\ 0 & 9 \\ 10 & -4 \end{bmatrix}.$$

The transpose of a matrix, for example $\boldsymbol{A}$ given in (16.1), is obtained by typing $\boldsymbol{A}'$ which gives

$$\boldsymbol{A}'=\begin{bmatrix} 5 & -5 & 9 \\ 2 & 0 & 10 \\ -1 & 9 & -4 \\ 7 & 7 & 10 \end{bmatrix}.$$

Additional matrix operations include

1. $inv(\boldsymbol{A})$ gives the inverse of the square matrix $\boldsymbol{A}$.
2. $pinv(\boldsymbol{A})$ gives the Moore–Penrose inverse of matrix $\boldsymbol{A}$.
3. $[\boldsymbol{A},\boldsymbol{B}]$ gives a horizontal concatenation of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ having the same number of rows.
4. $[\boldsymbol{A};\boldsymbol{B}]$ gives a vertical concatenation of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ having the same number of columns.
5. $det(\boldsymbol{A})$ gives the determinant of the square matrix $\boldsymbol{A}$.
6. $diag(\boldsymbol{A})$ gives the diagonal elements of the square matrix $\boldsymbol{A}$.
7. $min(\boldsymbol{A})$ gives the smallest element in each column of the matrix $\boldsymbol{A}$; $min(min(\boldsymbol{A}))$ gives the smallest element in the matrix $\boldsymbol{A}$.
8. $max(\boldsymbol{A})$ gives the largest element in each column of the matrix $\boldsymbol{A}$; $max(max(\boldsymbol{A}))$ gives the largest element in the matrix $\boldsymbol{A}$.
9. $mean(\boldsymbol{A})$ gives the average element in each column of the matrix $\boldsymbol{A}$; $mean(mean(\boldsymbol{A}))$ gives the average element in the matrix $\boldsymbol{A}$.
10. $median(\boldsymbol{A})$ gives the median of elements in each column of the matrix $\boldsymbol{A}$.
11. $std(\boldsymbol{A})$ gives the sample standard deviation in each column of the matrix $\boldsymbol{A}$.
12. $sum(\boldsymbol{A})$ gives the sum of elements in each column of the matrix $\boldsymbol{A}$; $sum(sum(\boldsymbol{A}))$ gives the sum of all elements in $\boldsymbol{A}$.
13. $prod(\boldsymbol{A})$ gives the product of elements in each column of the matrix $\boldsymbol{A}$. For example, if A=[2 3 7;5 4 9;1 2 10], then prod(A)=[10 24 630]. Note that $prod(1:n)$ gives $n!$.

14. $kron(\boldsymbol{A}, \boldsymbol{B})$ gives the direct product of the matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$.
15. $\boldsymbol{A} \backslash \boldsymbol{b}$ gives the solution to the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A}$ is an $n \times n$ nonsingular matrix and $\boldsymbol{b}$ is a column vector of n elements.
16. $rank(\boldsymbol{A})$ gives the rank of the matrix $\boldsymbol{A}$.
17. $randn(n)$ gives an $n \times n$ matrix of randomly distributed standard normal variables.
18. $size(\boldsymbol{A})$ gives the order of the matrix $\boldsymbol{A}$, that is, its numbers of rows and columns.
19. $trace(\boldsymbol{A})$ gives trace of the square matrix $\boldsymbol{A}$, that is, the sum of its diagonal elements.
20. $eig(\boldsymbol{A})$ gives the eigenvalues of the square matrix $\boldsymbol{A}$.
21. $[\boldsymbol{P}, \boldsymbol{D}] = eig(\boldsymbol{A})$ gives the diagonal matrix $\boldsymbol{D}$ of eigenvalues of the square matrix $\boldsymbol{A}$, and a matrix $\boldsymbol{P}$ whose columns are the corresponding eigenvectors. If $\boldsymbol{A}$ is symmetric, then this gives the spectral decomposition theorem such that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, where $\boldsymbol{P}$ is orthogonal (see Theorem 6.2).
22. $svd(\boldsymbol{A})$ gives a vector of non-negative elements. The positive elements are the square roots of the positive eigenvalues of $\boldsymbol{A}\boldsymbol{A}'$ (or, equivalently, of $\boldsymbol{A}'\boldsymbol{A}$), and are the singular values of the $m \times n$ matrix $\boldsymbol{A}$ $(m \leq n)$.
23. $[\boldsymbol{P}, \boldsymbol{R}, \boldsymbol{Q}] = svd(\boldsymbol{A})$ displays the matrix $\boldsymbol{R} = [\boldsymbol{D} \;\; \boldsymbol{0}]$, where the diagonal matrix $\boldsymbol{D}$ has non-negative elements, the positive ones of which are the square roots of the positive eigenvalues of $\boldsymbol{A}\boldsymbol{A}'$ (or, equivalently, of $\boldsymbol{A}'\boldsymbol{A}$), and are the singular values of the $m \times n$ matrix $\boldsymbol{A}$ $(m \leq n)$, as shown in formula (7.22). The two orthogonal matrices, $\boldsymbol{P}$ and $\boldsymbol{Q}$, are such that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{R}\boldsymbol{Q}'$ (this is the singular-value decomposition as described in Theorem 7.8).
24. $T = chol(\boldsymbol{A})$ gives the upper triangular matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}$, where $\boldsymbol{A}$ is positive definite (see the Cholesky decomposition theorem 6.3).

Some elementary matrices are

1. eye(n) gives an identity matrix of order $n \times n$.
2. zeros(m,n) gives a matrix of zeros of order $m \times n$.
3. ones(m,n) gives a matrix of ones of order $m \times n$.

Example 16.1 *Consider the matrix* $\boldsymbol{A}$ *given in (16.1) and its* 3×3 *submatrix*

$$\boldsymbol{B} = \begin{bmatrix} 5 & 2 & -1 \\ -5 & 0 & 9 \\ 9 & 10 & -4 \end{bmatrix}. \tag{16.2}$$

The rank of $\boldsymbol{A}$ *is* $rank(\boldsymbol{A}) = 3$*. the determinant of* $\boldsymbol{B}$ *is* $det(B) = -278$*. Its inverse is the matrix*

$$inv(\boldsymbol{B}) = \begin{bmatrix} 0.3237 & 0.0072 & -0.0647 \\ -0.2194 & 0.0396 & 0.1439 \\ 0.1799 & 0.1151 & -0.0360 \end{bmatrix}$$

The trace of $\boldsymbol{B}$ *is trace(*$\boldsymbol{B}$*)* $= 1$. *The eigenvalues and corresponding eigenvectors of* $\boldsymbol{B}$ *are given by* $[\boldsymbol{V},\boldsymbol{D}] = eig(\boldsymbol{B})$, *where* $\boldsymbol{D}$ *is a diagonal matrix of eigenvalues and* $\boldsymbol{V}$ *is a matrix whose columns are the corresponding eigenvectors of* $\boldsymbol{B}$,

$$\boldsymbol{D} = \begin{bmatrix} -10.3659 & 0 & 0 \\ 0 & 3.3426 & 0 \\ 0 & 0 & 8.0234 \end{bmatrix}, \qquad \boldsymbol{V} = \begin{bmatrix} 0.1305 & -0.8031 & 0.2075 \\ -0.6133 & 0.5434 & 0.6701 \\ 0.7790 & -0.2444 & 0.7127 \end{bmatrix}.$$

Example 16.2 *Consider the matrices*

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 7 \\ 8 & 10 & 9 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 1 & 5 & 6 \\ 5 & 4 & 2 \\ 6 & 2 & 8 \end{bmatrix}.$$

The singular-value decomposition of $\boldsymbol{A}$ *is* $\boldsymbol{A} = \boldsymbol{P}[\boldsymbol{D}\ \ \boldsymbol{0}]\boldsymbol{Q}'$, *where* $\boldsymbol{D}$ *is a diagonal matrix and* $\boldsymbol{P}$ *and* $\boldsymbol{Q}$ *are orthogonal matrices given by*

$$\boldsymbol{P} = \begin{bmatrix} -0.3078 & -0.9515 \\ -0.9515 & 0.3078 \end{bmatrix}, \qquad [\boldsymbol{D}\ \ \boldsymbol{0}] = \boldsymbol{R} = \begin{bmatrix} 16.3499 & 0 & 0 \\ 0 & 5.6286 & 0 \end{bmatrix},$$

$$\boldsymbol{Q} = \begin{bmatrix} -0.5032 & 0.0993 & -0.8584 \\ -0.5631 & 0.7158 & 0.4129 \\ -0.6555 & -0.6912 & 0.3043 \end{bmatrix}.$$

The spectral decomposition of the symmetric matrix $\boldsymbol{B}$ *is* $\boldsymbol{B} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}'$, *where* $\boldsymbol{\Lambda}$ *is a diagonal matrix of eigenvalues of* $\boldsymbol{B}$ *and* $\boldsymbol{S}$ *is an orthogonal matrix whose columns are the corresponding eigenvectors of* $\boldsymbol{B}$ *and are given by*

$$\boldsymbol{\Lambda} = \begin{bmatrix} -4.0604 & 0 & 0 \\ 0 & 3.5812 & 0 \\ 0 & 0 & 13.4792 \end{bmatrix}, \qquad \boldsymbol{S} = \begin{bmatrix} -0.8339 & -0.1686 & 0.5255 \\ 0.4321 & -0.7917 & 0.4318 \\ 0.3432 & 0.5872 & 0.7331 \end{bmatrix}.$$

Example 16.3 *Consider the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 27 & -11 & -5 \\ -11 & 109 & 128 \\ -5 & 128 & 154 \end{bmatrix}.$$

This matrix is positive definite as its eigenvalues are all positive given by 0.4956, 27.5358, 261.9686. These eigenvalues can be obtained by using the matrix operation eig($\boldsymbol{A}$). Since $\boldsymbol{A}$ is positive definite, its Cholesky decomposition can be done by using the matrix operation chol($\boldsymbol{A}$). This results in the upper triangular matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}$, where $\boldsymbol{T}$ is given by

$$\boldsymbol{T} = \begin{bmatrix} 5.1962 & -2.1170 & -0.9623 \\ 0 & 10.2234 & 12.3210 \\ 0 & 0 & 1.1255 \end{bmatrix}.$$

Example 16.4 *Consider the symmetric matrix*

$$\boldsymbol{A} = \begin{bmatrix} 3 & -5 & 10 & 7 \\ -5 & 15 & 0 & 9 \\ 10 & 0 & -1 & 12 \\ 7 & 9 & 12 & 4 \end{bmatrix}$$

The diagonal elements of $\boldsymbol{A}$ can be obtained by using the matrix operation diag($\boldsymbol{A}$), which gives the vector $\boldsymbol{b} = (3, 15, -1, 4)'$. The trace of $\boldsymbol{A}$ is obtained by using trace($\boldsymbol{A}$), which gives the value tr($\boldsymbol{A}$) $= 21$. The spectral decomposition of $\boldsymbol{A}$ can be done by using the matrix operation $[\boldsymbol{P}, \boldsymbol{D}] =$ eig($\boldsymbol{A}$), which results in $\boldsymbol{A} = \boldsymbol{PDP}'$, where $\boldsymbol{D}$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix whose columns are the corresponding eigenvectors. These are given by

$$\boldsymbol{D} = \begin{bmatrix} -12.3366 & 0 & 0 & 0 \\ 0 & -7.3400 & 0 & 0 \\ 0 & 0 & 17.5420 & 0 \\ 0 & 0 & 0 & 23.1347 \end{bmatrix},$$

$$\boldsymbol{P} = \begin{bmatrix} 0.1982 & 0.7488 & -0.5438 & -0.3229 \\ -0.1522 & 0.3670 & 0.7574 & -0.5181 \\ -0.7809 & -0.2443 & -0.3504 & -0.4558 \\ 0.5725 & -0.4949 & -0.0882 & -0.6477 \end{bmatrix}.$$

Example 16.5 *Consider the two matrices*

$$\boldsymbol{A} = \begin{bmatrix} 5 & 2 & -1 \\ 7 & 11 & 9 \\ 1 & 14 & 8 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 0.5 & 0.20 & 0.3 \\ 12.0 & 1.0 & 7.9 \\ -1.6 & 6.9 & 8.5 \end{bmatrix}.$$

The product of $\mathbf{A}$ *and* $\mathbf{B}$ *is*

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} 28.1 & -3.9 & 8.8 \\ 121.1 & 74.5 & 165.5 \\ 155.7 & 69.4 & 178.9 \end{bmatrix}.$$

The element-by-element product (array product) of $\mathbf{A}$ *and* $\mathbf{B}$ *is*

$$\mathbf{A}.*\mathbf{B} = \begin{bmatrix} 2.5 & 0.4 & -0.3 \\ 84.0 & 11.0 & 71.1 \\ -1.6 & 96.6 & 68.0 \end{bmatrix}.$$

The quotient of $\mathbf{A}$ *and* $\mathbf{B}$, $\mathbf{A}/\mathbf{B} = \mathbf{A} * inv(\mathbf{B})$, *is*

$$\mathbf{A}/\mathbf{B} = \begin{bmatrix} 26.3780 & -0.7316 & -0.3687 \\ 39.9510 & -1.0037 & 0.5817 \\ 47.3376 & -1.7673 & 0.9130 \end{bmatrix}.$$

The element-by-element quotient of the elements $\mathbf{A}$ *by those of* $\mathbf{B}$, *namely* $\mathbf{A}./\mathbf{B}$, *is*

$$\mathbf{A}./\mathbf{B} = \begin{bmatrix} 10.0 & 10.0 & -3.3333 \\ 0.5833 & 11.0 & 1.1392 \\ -0.625 & 2.029 & 0.9412 \end{bmatrix}.$$

The sum of squares of the elements of $\mathbf{A}$ *is equal to the trace of the matrix* $\mathbf{A} * \mathbf{A}'$, *that is,* $tr(\mathbf{A} * \mathbf{A}') = 542$. *The square root,* $\sqrt{(542)} = 23.281$, *of this trace is called the* Frobenius norm.

Example 16.6 *Consider again the two matrices,* $\mathbf{A}$ *and* $\mathbf{B}$, *in Example 16.5. Let* $\mathbf{b}$ *be the column vector* $(3, -2, 8)'$. *The solution of the linear system* $\mathbf{Ax} = \mathbf{b}$ *is obtained by using* $\mathbf{A}\backslash\mathbf{b}$. *The solution is* $(-0.4798, 1.7224, -1.9542)'$. *The matrix* $\mathbf{A}.\hat{}3 + \mathbf{B}.\hat{}2$ *is equal to*

$$\mathbf{A}.\hat{}3 + \mathbf{B}.\hat{}2 = \begin{bmatrix} 125.3 & 8.0 & -0.9 \\ 487.0 & 1332.0 & 791.4 \\ 3.6 & 2791.6 & 584.3 \end{bmatrix}$$

The smallest element in $\mathbf{A}$ *is* $min(min(\mathbf{A})) = -1$. *The maximum elements in the three columns of* $\mathbf{B}$ *are* $max(\mathbf{B}) = (12.0, 6.9, 8.5)'$. *The average elements in the three columns* *of* $\boldsymbol{A}$ *are mean*$(\boldsymbol{A}) = (4.3333, 9.0000, 5.3333)'$. *The sample standard deviations in the three columns of* $\boldsymbol{B}$ *are std*$(\boldsymbol{B}) = (7.3214, 3.6592, 4.5709)'$.

16.4 TWO- AND THREE-DIMENSIONAL PLOTS

MATLAB has an excellent capability to produce two-dimensional as well as three-dimensional plots. These plots can be easily obtained with very few commands as we now demonstrate.

For two-dimensional plots, we can either have

(a) a vector of x-coordinates, $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, and a vector of y-coordinates, $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$, that have identical array forms, or
(b) a function $y = f(x)$, where $f(x)$ can be evaluated for a given set of x values within a certain specified domain.

In case of (a), the points (x_i, y_i), $i = 1, 2, \dots, n$, are formed and joined together by straight lines. This can be established by using the command *plot*(x, y). In case of (b), a vector of x values is created within a specified interval and the function $f(x)$ is evaluated at each of the x values. This process is followed by plotting the resulting values. The plot is obtained by again using the command *plot*(x, y).

To demonstrate part (a), suppose we are given the two vectors, $\boldsymbol{x} = (1\ \ 3\ \ 5\ \ 6\ \ 8\ \ 9\ \ 10)'$ and $\boldsymbol{y} = (3\ \ 8\ \ 13\ \ 12\ \ 8\ \ 10\ \ 12)'$. To plot these vectors we use the following MATLAB statements:

$\boldsymbol{x} = [1\ \ 3\ \ 5\ \ 6\ \ 8\ \ 9\ \ 10]$;
$\boldsymbol{y} = [3\ \ 8\ \ 13\ \ 12\ \ 8\ \ 10\ \ 12]$;
plot$(\boldsymbol{x}, \boldsymbol{y})$

Note that we used semicolons to produce new rows, that is, to separate them. The result of the plot statement is shown in Figure 16.1.

As an example of part (b), consider plotting the function $y = cos(x)$ over the interval $[0, 2\pi]$ which is partitioned with mesh size $\pi/100$. The following MATLAB statements are needed:

$x = 0 : pi/100 : 2 * pi$;
$y = cos(x)$;
plot(x,y)

The resulting plot is shown in Figure 16.2.

Additional MATLAB commands can be added to a given graph. These include the *title* command, the *xlabel*, *ylabel*, and *legend* commands. The *title* command adds a title to the graph. The *xlabel* and *ylabel* add labels to the x-axis and the y-axis, respectively. The *legend* command places an explanation displayed in the figure that gives added information about the figure. For example, we can add a title command and xlabel and ylabel commands to Figure 16.2. The MATLAB statements needed to do this are

$x = 0 : pi/100 : 2 * pi$;
$y = cos(x)$;
plot(x,y); xlabel('x'); ylabel('y'); title('Plot of Cosine Function')

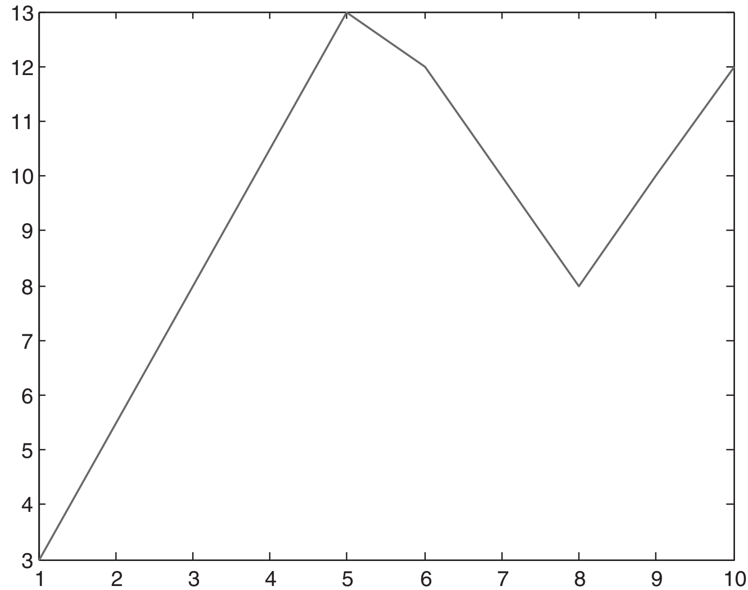


Figure 16.1 *Plot of* **x** *and* **y** *Vectors.*

Note that the xlabel, ylabel, and title commands should be placed on the same line after the plot command.

Several functions can be plotted in the same graph. In this case, the different plots can be identified by using different markers. For example, consider plotting the functions $y1 = x^2 - 1$ and $y2 = 2x + 1$. The MATLAB statements for such a plot are

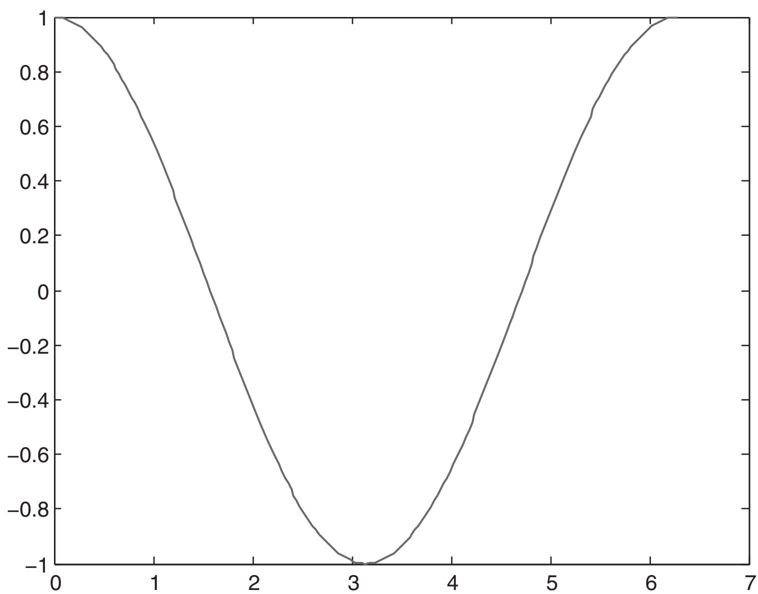


Figure 16.2 *Plot of Cosine Function.*

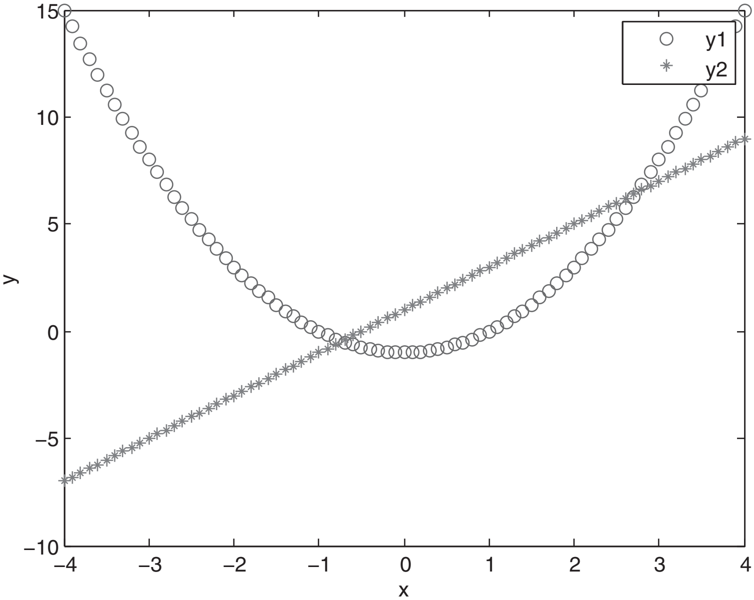


Figure 16.3 *Plot of Two Polynomial Functions.*

$x = 0 : pi/15 : 2 * pi;$
$y1 = x^2 - 1;$
$y2 = 2 * x + 1;$
plot(x,y1,'o',x,y2,'*'); hold on; title('Two Polynomial Functions'); xlabel('x-axis'); ylabel('y-axis'); legend('y1','y2')

The resulting plot is shown in Figure 16.3. The use of the hold on command will be explained later in the next section.

The above commands are also available in the *Insert* menu in the Figure window. Thus by clicking, for example, on *x label*, *y label*, *Title*, and *Legend* we can insert labels for the *x* and *y* axes, type a title, and include a legend for the figure. By clicking on *Insert* and *TextBox* followed by clicking on the figure, we get a box inside of which we can include additional information about the figure. If we then click somewhere else in the figure we can finish entering the text.

Let us now reconsider earlier plots used in Chapter 2. These concern plotting of the 4×4 matrix $\boldsymbol{M}$ in (2.8), which is shown in Figure 2.1, and the 4×4 identity matrix in (2.9), which is shown in Figure 2.2. The needed MATLAB statements to generate Figure 2.1 are

$\boldsymbol{M} = [15\ \ 2\ \ 4\ \ 14; 5\ \ 10\ \ 11\ \ 7; 8\ \ 8\ \ 6\ \ 13; 5\ \ 14\ \ 13\ \ 1];$
$\boldsymbol{m1} = \boldsymbol{M}(1:4, 1);$
$\boldsymbol{m2} = \boldsymbol{M}(1:4, 2);$
$\boldsymbol{m3} = \boldsymbol{M}(1:4, 3);$
$\boldsymbol{m4} = \boldsymbol{M}(1:4, 4);$
$\boldsymbol{d} = [1\ ; 2\ ; 3\ ; 4];$
plot[**d**, **m1**, '+-', **d** , **m2**, 'o–', **d** , **m3**, , 'k*', **d** , **m4**, 'd-.'); xlabel('row number'); ylabel('column element value'); title('Plot of the Matrix **M**'); legend('col1','col2','col3','col4')

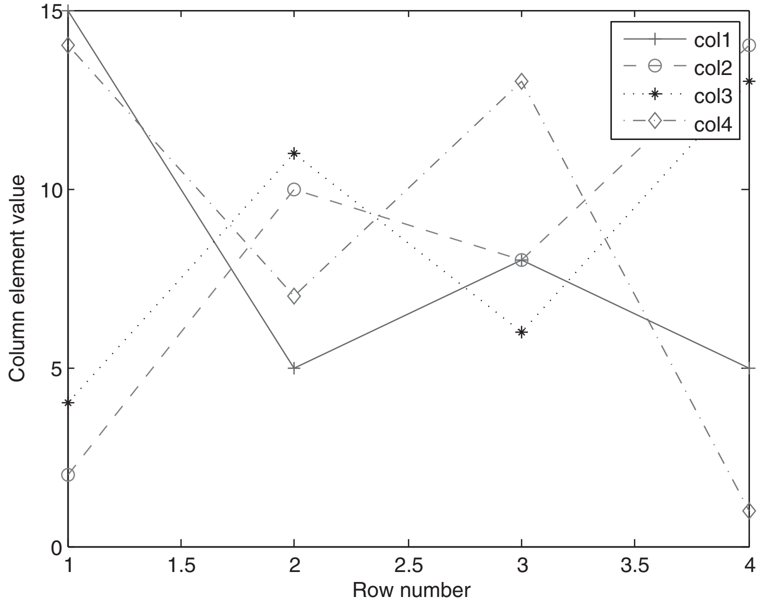


Figure 16.4 *Plot of the Matrix M.*

For convenience, the plot of the Matrix $\boldsymbol{M}$ is displayed again in Figure 16.4. A similar procedure can be used to plot the 4×4 identity matrix shown in Figure 2.2. It should be noted that in Figure 16.4 each column of the matrix $\boldsymbol{M}$ is represented by a separate broken line and the four lines are marked by different markers.

Let us also consider another earlier plot used in Section 6.1.1 and shown in Figure 6.2. It concerns the plotting of the eigenvalues of a random matrix of order 50×50. The elements of this matrix were randomly selected from the standard normal distribution. The MATLAB statements needed to produce Figure 6.2 are

$\boldsymbol{A} = randn(50);$
$e = eig(\boldsymbol{A});$
plot(real(e),imag(e),'k*'); xlabel('real'); ylabel('imaginary'); title('Plot of Eigenvalues of a Matrix of Order 50 × 50')

For convenience, the plot of eigenvalues is shown again in Figure 16.5.

16.4.1 Three-Dimensional Plots

There are four commands available in MATLAB for three-dimensional plots. These commands are *plot3*, *mesh*, *surf*, and *contour*.

1. *plot3* is used to plot curves in a three-dimensional space. For example, consider the following plot:

$t = 0 : 0.01 : 18 * pi;$
$x = sin(t);$
$y = cos(2 * t);$
$z = t.^5;$
plot3(x,y,z)

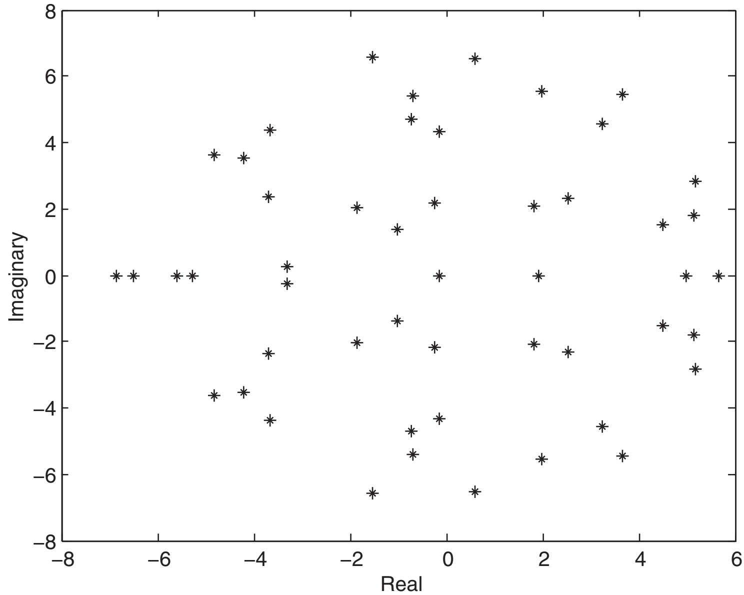


Figure 16.5 *Plot of Eigenvalues of a* 50×50 *Matrix.*

Note that the use of the dot after t in $t.^5$ will be explained later. The resulting plot is shown in Figure 16.6.

2. *mesh* is used to create a three-dimensional plot of the elements of a matrix. The values of the elements of the matrix are marked along the z-axis and the numbers of the corresponding rows and columns are marked along the x- and y-axes. For example, we can have the following MATLAB statements concerning a 4×4 matrix $\boldsymbol{A}$:

A = [50 30 9 25; 100 60 18 55; 2 10 18 75; 52 49 19 100]
mesh(**A**)

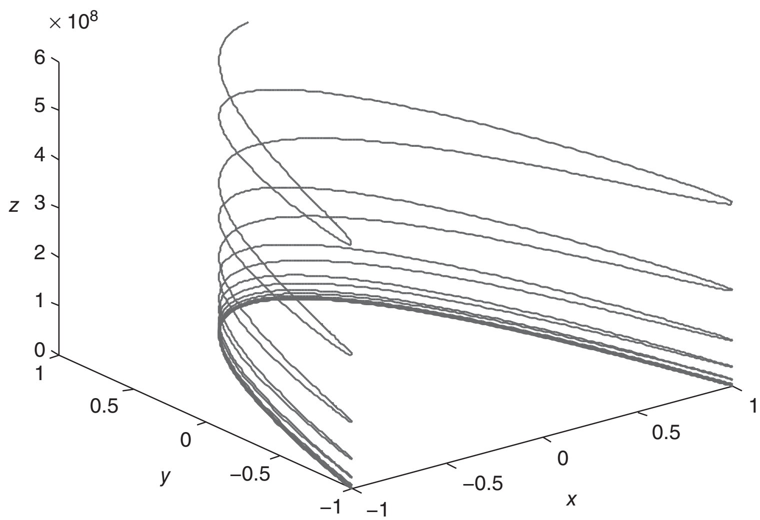


Figure 16.6 *Curves in a Three-Dimensional Space.*

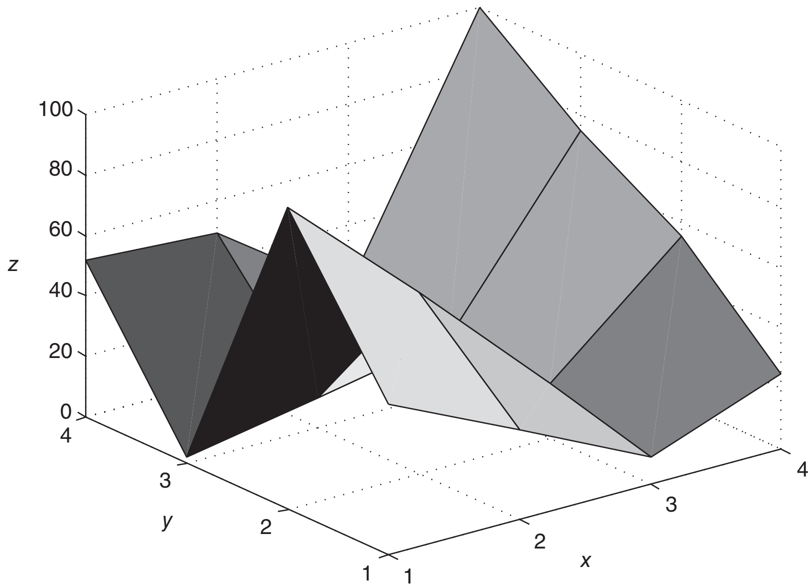


Figure 16.7 *Surf Applied to Matrix* **A**.

3. *surf* is similar to the *mesh*. Applying it to the matrix $\boldsymbol{A}$ of the previous example and using the statements

A = [50 30 9 25; 100 60 18 55; 2 10 18 75; 52 49 19 100]
surf(**A**)
we get the graph shown in Figure 16.7.

It is also possible to draw a three-dimensional graph of a function $z = f(x, y)$ of two variables, x and y, over a rectangular region A in a two-dimensional space. The sides of the region A are partitioned by defining the vectors x and y using, for example, the command, $[x, y] =$ meshgrid$(-2 : .20 : 2, -2 : .20 : 2)$. A variable z can then be computed by evaluating the function $f(x, y)$, using mesh or surf, over the x, y plane.

Example 16.7 *Consider plotting the three-dimensional graph of the function z given by*

$$s = \sqrt{3 * (x.^2) + 9 * (y.^2)} + 1.0,$$
$$z = sin(s)./(s.(0.25)), \qquad (16.3)$$

by using the MATLAB statements,

$[x, y] = meshgrid(-9.0 : .20 : 9, -9.0 : .20 : 9.0)$;
$s = sqrt(3 * (x.^2) + 9 * (y.^2)) +$ *1.0;*
$z = sin(s)./(s.(0.25))$;
mesh(z) xlabel('x'); ylabel('y'); zlabel('z'); title('The Graph of the Function z')

The plot of this function is shown in Figure (16.8).

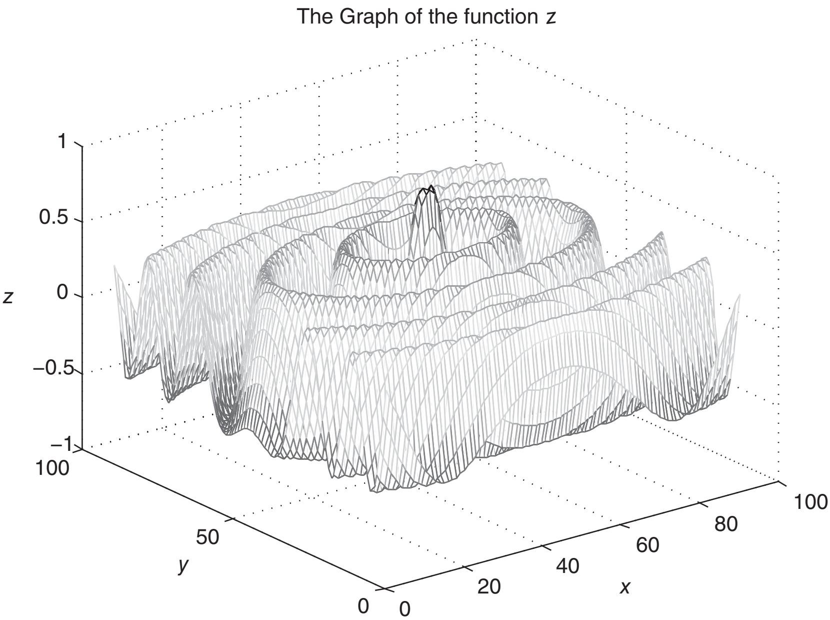


Figure 16.8 *A Mesh Plot.*

Note that in this example as well as in two previous examples, the evaluation of a function that includes products, divisions, or exponentiations of variables, each operation must be preceded by a period to ensure that it is performed element by element, that is, it is done on an entry-wise basis. For example, if we have an expression such as x^2y^2, it should be written in MATLAB as $(x.^2). * y.^2$. Leaving out, for example, the dot before the multiplication operation as in $(x.^2) * y.^2$, will lead to an error message. See, for example, Davis and Sigmon (2005, pp. 71 and 79) and Kattan (2008, pp. 71–72).

4. The fourth command for plotting in a three-dimensional space is *contour* which produces contour plot of a surface by dropping cross sections of the surface parallel to the x-y plane on that plane. This command is activated by using *contour* in place of *mesh* or *surf*. For example, let us consider applying this command to the matrix $\boldsymbol{A}$ used earlier in this section. The MATLAB statements needed to generate such a plot are

A=[50 30 9 25; 100 60 18 55; 2 10 18 75; 52 49 19 100]
contour(**A**)

It is also possible to combine the commands *mesh* or *surf* with the command *contour*. This can be done by using the commands *meshc* and *surfc* instead of *mesh* and *surf*, respectively. For example, applying *meshc* to the matrix $\boldsymbol{A}$ and using the statements

A=[50 30 9 25; 100 60 18 55; 2 10 18 75; 52 49 19 100]
meshc(**A**)

we get the plot shown in Figure 16.9.

Another MATLAB command that can be useful with regard to plotting is the *hold on* command. It is used to add additional plots to an existing one. This is needed because the use of a new plot will erase the previous plot if this command is not used. Thus, *hold on*

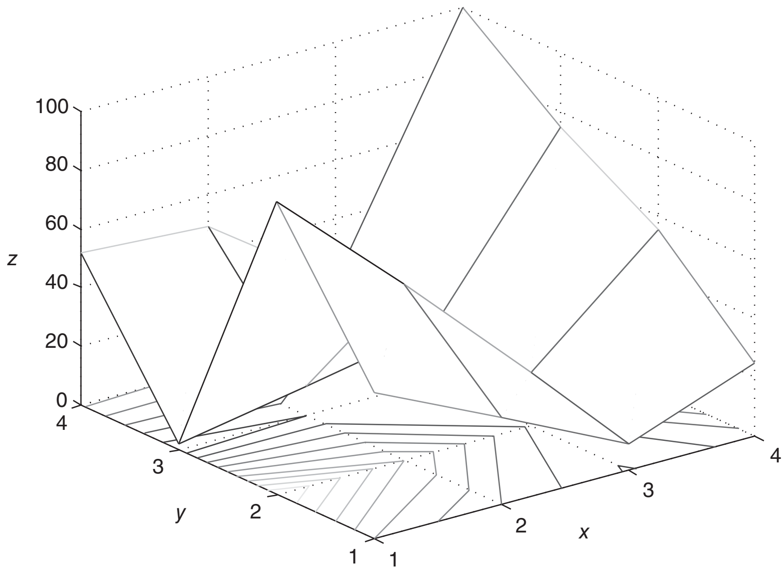


Figure 16.9 Mesh-Contour of Matrix A.

retains the existing plot and certain axes properties so that additional plots can be added to the current plot. The subsequent use of the command *hold off* cancels the effect of *hold on* so that new plots delete existing plots, and the axes properties are reset to their original settings before the introduction of the new plots. For example, the following statements can be tried:

$x = (-10 : 0.1 : 10);$
$y = x.^2;$
$z = x.^3;$
$w = cos(x);$
plot(x, y); hold on; plot(x, z); plot(x, w)

The first statement gives a range of values with an increment of 0.1 for the variable x. The remaining statements introduce a plot of x versus y followed by the addition of two new plots of x versus z and x versus w.

There are several useful references on MATLAB that can be helpful to the reader, which were also helpful to the writing of this chapter, including Davis and Sigmon (2005), Hiebeler (2015), Kattan (2008), López (2014), and Sizemore and Mueller (2015), among others.

REFERENCES

Davis, T. A. and Sigmon, K. (2005). *MATLAB Primer*, 7th ed. Chapman & Hall/CRC, Boca Raton, FL.

Hiebeler, D. E. (2015). *R and MATLAB*. Chapman & Hall/CRC, Boca Raton.

Kattan, P. I. (2008). *MATLAB for Beginners*. Author.

López, C. P. (2014). *MATLAB Matrix Algebra*. Petra Books.

Sizemore, J. and Mueller, J. P. (2015). *MATLAB for Dummies*. John Wiley & Sons, Hoboken, NJ.

EXERCISES

Do the following exercises using MATLAB.

16.1 Consider the matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \begin{bmatrix} 5 & 2 & 1 \\ 9 & 6 & 5 \\ 3 & 10 & 15 \end{bmatrix}.$$

(a) Find the rank of $\boldsymbol{A}$ and demonstrate it is nonsingular.

(b) Find the determinant of $\boldsymbol{A}$.

(c) Find the eigenvalues of $\boldsymbol{A}$.

(d) Find the trace of $\boldsymbol{A}$.

16.2 Let $\boldsymbol{A}$ be the matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}$$

(a) Find the eigenvalues of $\boldsymbol{A}$ and conclude that it is positive definite.

(b) Verify that $|\boldsymbol{A}| \leq \prod_{i=1}^{3} a_{ii}$, where a_{ii} is the ith diagonal element of $\boldsymbol{A}$ $(i = 1, 2, 3)$.

16.3 Consider again the matrix in Exercise 2.

(a) Verify that its leading principal minors are all positive.

(b) Verify that $\boldsymbol{A}'\boldsymbol{A}$ is positive definite.

16.4 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$$

(a) Verify that $\boldsymbol{A}$ is positive semidefinite.

(b) Find a matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}^2$.

16.5 Consider once more the matrix $\boldsymbol{A}$ in Exercise 2.

(a) Find the inverse of $\boldsymbol{A}$.

(b) Verify that $tr(\boldsymbol{A})\, tr(\boldsymbol{A}^{-1}) \geq 9$.

16.6 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 4 & -2 \\ -4 & 0 & -8 \\ 2 & 8 & 0 \end{bmatrix}$$

(a) Show that $\boldsymbol{A}$ is skew-symmetric.

(b) Verify that $(\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} + \boldsymbol{A})^{-1}$ is orthogonal.

16.7 Consider the positive definite matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

(a) Find the Cholesky decomposition of $\boldsymbol{A}$.

(b) Apply the spectral decomposition theorem to $\boldsymbol{A}$.

16.8 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 7 & 12 \\ 8 & 5 \\ 10 & 9 \end{bmatrix}.$$

(a) Find the Moore–Penrose inverse of $\boldsymbol{A}$. Denote this inverse by $\boldsymbol{B}$.

(b) Verify that $\boldsymbol{ABA} = \boldsymbol{A}$, and $\boldsymbol{BAB} = \boldsymbol{B}$.

16.9 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix}.$$

(a) Show that $\boldsymbol{BA}$ is idempotent, where $\boldsymbol{B}$ is the Moore–Penrose inverse of $\boldsymbol{A}$.

(b) Find the eigenvalues of $\boldsymbol{BA}$.

(c) Find the trace of $\boldsymbol{BA}$.

(d) Find the rank of $\boldsymbol{BA}$.

16.10 Consider the following matrices:

$$\boldsymbol{A} = \begin{bmatrix} 3 & 5 & 8 \\ 6 & 1 & 2 \\ 9 & 4 & 3 \\ 2 & 0 & 10 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 4 & 6 & -3 \\ -5 & 6 & 2 \\ 2 & 10 & 4 \\ 8 & -5 & 2 \end{bmatrix}.$$

Verify that

$$[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \leq tr(\boldsymbol{A}'\boldsymbol{A})\, tr(\boldsymbol{B}'\boldsymbol{B}),$$

which is the Cauchy–Schwarz inequality for matrices.

16.11 Consider the matrix

$$A = \begin{bmatrix} 10 & 2 & 6 \\ -5 & -11 & -8 \end{bmatrix}.$$

(a) Find the singular-value decomposition of $\boldsymbol{A}$.

(b) List the singular values of $\boldsymbol{A}$ and verify their equality to the square roots of the positive eigenvalues of $\boldsymbol{AA'}$.

16.12 Consider the two matrices

$$\boldsymbol{A} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 3 & 4 \end{bmatrix}.$$

(a) Find the direct product of $\boldsymbol{A}$ and $\boldsymbol{B}$.

(b) Verify that $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.

16.13 Consider the two matrices

$$\boldsymbol{A} = \begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}.$$

(a) Verify that $tr[(\boldsymbol{AB})^2] \leq tr(\boldsymbol{A}^2\boldsymbol{B}^2)$.

(b) Verify that $|\boldsymbol{A}'\boldsymbol{B}|^2 \leq |\boldsymbol{A}'\boldsymbol{A}|\ |\boldsymbol{B}'\boldsymbol{B}|$.

16.14 Plot the function,

$$y = sin(x)\, cos(x) + cos(3\, x),$$

over the interval $[0, 2\pi]$.

16.15 Plot the function,

$$z = y\ exp(-3\, x^2 - 7\, y^2),$$

using surf(z) over the range [-2,2] for each of x and y.

17

Use of R in Matrix Computations

R is a computer program with emphasis on statistical applications. It differs from MATLAB in the sense that the latter is more oriented toward applications in applied mathematics. R performs computational procedures and matrix calculations, and, like MATLAB, has efficient graphical capabilities. A very attractive feature of R is its being available for free without any cost to the user. It is also an open source system whereby individual users can contribute to the development of its software. R is based on the commercial S-PLUS package and its earlier version the S package which was developed by John Chambers and several of his colleagues at Bell Laboratories. Chambers was also helpful in the creation and development of R. It is believed that the choice of the letter R was based on the first initials of Ross Ihaka and Robert Gentleman at the University of Auckland in New Zealand who wrote the first version of R and used it for teaching purposes. R has become very popular among many users. It is often described as "the most widely used programming language for data science."

Installation of R

Before beginning to use R it must be installed on one's computer system. To do so, the following steps can be taken:

(a) Go to the website www.r-project.org
(b) Click on CRAN on the left side on the screen under **DOWNLOAD**.
(c) Select a CRAN Mirror Site in the USA
(d) Under **Download and Install R**, select your operating system (e.g., Windows).
(e) Click on the link labelled **base**.
(f) Click on the link for the current version of R (currently it is R.3.2.3 for Windows).

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

After installation, R can be started by double clicking on the R icon. To exit R, type q() in the workspace. Note that R is case sensitive.

Some Basic Concepts

1. **Objects**. Anything that R works on is called an object. This includes vectors, matrices, and functions. For example,

$$\mathbf{x} = \mathrm{c}(10, 15, 20, 36)$$

creates a vector of length 4 using the c() function in order to concatenate the selected elements. There are also several arithmetic/matrix operators that can be used for computational purposes. These include

Addition + (e.g., **A** + **C**)

Subtraction - (e.g., **A** - **B**)

Multiplication %*% (e.g., **A**%*%**B**). This is a standard multiplication and can be used for matrices.

Division / (e.g., **A/B**)

Exponentiation ^ (e.g., a^n produces a raised to the nth power).

Element-by-element multiplication * (e.g., **A*****B**, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of the same order).

Kronecker (direct) product %×% (e.g., **A**%×%**B**, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices).

Direct sum of matrices direct.sum (e.g., the direct sum of $\boldsymbol{A}$ and $\boldsymbol{B}$ is direct.sum(**A,B**)).

Transpose of a matrix t() (e.g., t(**A**)).

2. **Diagonal Matrices**. These can be set up by either creating one from a list such as

$$\mathbf{a} = \mathrm{c}(5, 10, 15, 25)$$

$\boldsymbol{A} = \mathrm{diag}(\mathbf{a})$, which gives

$$\boldsymbol{A} = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 15 & 0 \\ 0 & 0 & 0 & 15 \end{bmatrix},$$

or can be extracted from a matrix. For example, if $\boldsymbol{B}$ is the matrix

$$\boldsymbol{B} = \begin{bmatrix} 7 & -1 & 4 & 12 \\ -2 & 8 & 17 & 25 \\ 9 & 11 & 19 & 21 \\ -5 & -9 & 11 & 34 \end{bmatrix},$$

then diag(**B**) gives the diagonal elements of $\boldsymbol{B}$, namely, 7 8 19 34.

The trace of $\boldsymbol{B}$ in R is denoted by $sum(diag(\boldsymbol{A}))$.

The identity matrix of order $n \times n$ is denoted by $diag(n)$.

3. **Constructing a Matrix**. Matrices can be created by using the function matrix(). For example,

$$\boldsymbol{A} = \text{matrix}(\mathbf{c}(2,6,9,11,18,24), \text{nrow} = 3, \text{ncol} = 2)$$

creates a 3×2 matrix where the elements are entered column by column and we get

$$\boldsymbol{A} = \begin{bmatrix} 2 & 11 \\ 6 & 18 \\ 9 & 24 \end{bmatrix}.$$

It is possible to delete the terms "nrow=" and "ncol=" and just use their values. In this case, it is assumed that they are in the order row and column. Hence, using the statement

$$\boldsymbol{A} = \text{matrix}(\mathbf{c}(2,6,9,11,18,24), 3,\ 2)$$

should lead to the same result.

To point to a particular element of a matrix, for example, the element in row 1 and column 2 of $\boldsymbol{A}$, we use **A**[1,2], which gives the value $b = 11$, if we were to type $b = \boldsymbol{A}[1,2]$. To point to a particular column of a matrix, for example, column 1 of $\boldsymbol{A}$, we use **A**[,1], which gives the elements 2 6 9. Likewise, to point to a particular row of a matrix, for example, row 2 of $\boldsymbol{A}$, we use **A**[2,], which gives the elements 6 18. Furthermore, to select a submatrix of a matrix, for example, the submatrix of $\boldsymbol{A}$ consisting of the last two rows and the two columns of $\boldsymbol{A}$, we use **A**[2:3,1:2], which gives the submatrix of the stated elements. It is also possible to select a submatrix when the selected rows or columns are not consecutive. For example, selecting a submatrix of $\boldsymbol{A}$ consisting of rows 1 and 3 of $\boldsymbol{A}$ and columns 1 and 2, we use A[c(1,3), c(1,2)],which gives the stated submatrix.

To join two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ side by side, use *cbind*($\boldsymbol{A}, \boldsymbol{B}$). To stack them on top of each other, use *rbind*($\boldsymbol{A}, \boldsymbol{B}$).

The order of a matrix $\boldsymbol{A}$ is denoted by *dim*($\boldsymbol{A}$).

An $m \times n$ matrix consisting of elements equal to p is denoted in R by matrix(p,m,n). For example, matrix(1,4,4) gives the matrix of ones of order 4×4.

4. **Eigenvalues and Eigenvectors**. The eigenvalues and eigenvectors of a matrix $\boldsymbol{A}$ can be obtained by using eigen(**A**). For example, let $\boldsymbol{A}$ be the matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 12 & 18 & 6 \\ 12 & 2 & 3 & 9 \\ 18 & 3 & 10 & 11 \\ 6 & 9 & 11 & 12 \end{bmatrix}. \tag{17.1}$$

Using eigen(**A**) produces the eigenvalues 37.5633128, 5.3604021, 0.9911004, −15.9148152, and the eigenvectors

$$\begin{array}{rrrr} -0.5300973 & 0.3926523 & 0.3083266 & -0.6853872 \\ -0.3549924 & -0.2455179 & 0.7649733 & 0.4780348 \\ -0.5847639 & 0.4482265 & -0.4453918 & 0.5086947 \\ -0.5010273 & -0.7646163 & -0.3483917 & -0.2072602 \end{array}$$

5. **Spectral Decomposition Theorem**. Let $\boldsymbol{A}$ be a symmetric matrix, for example, the matrix $\boldsymbol{A}$ given in (17.1). Then, the spectral decomposition of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{P\Lambda P'}$, where $\boldsymbol{P}$ is the matrix of eigenvectors seen earlier and $\boldsymbol{\Lambda}$ is the diagonal matrix of corresponding eigenvalues. Example 17.4 (to be presented later) shows how the spectral decomposition can be carried out in R using the eigen(**A**) command.
6. **The Singular-Value Decomposition**. Let $\boldsymbol{A}$ be a matrix of order $m \times n$ and rank r, then $\boldsymbol{A}$ has the singular-value decomposition as shown in expression (7.22). Using the details given in the proof of Theorem 7.8, it is easy to show that this expression can also be written as

$$\boldsymbol{A} = \boldsymbol{P}_1 \boldsymbol{D}_1 \boldsymbol{Q}_1', \tag{17.2}$$

where $\boldsymbol{P}_1$ is of order $m \times r$ whose columns are eigenvectors of $\boldsymbol{A'A}$ corresponding to the positive eigenvalues of $\boldsymbol{AA'}$, which are the diagonal elements of the diagonal matrix $\boldsymbol{D}_1^2$, and $\boldsymbol{Q}_1$ is the matrix defined in expression (7.24). Using R, the singular-value decomposition of $\boldsymbol{A}$ can be invoked by using the command svd(**A**). This produces the diagonal elements of $\boldsymbol{D}_1$ (that is, the singular values) and the columns of the matrices $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$. For example, consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 6 & 11 & -1 & 5 \\ 5 & 0 & 2 & 1 \\ 10 & 15 & -5 & 7 \end{bmatrix}. \tag{17.3}$$

The singular values and the columns of $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$ are
Singular values:
24.147932 4.910129 2.183576

	−0.5551295	0.007788482	0.83172747
$\boldsymbol{P}_1$	−0.1032498	−0.992867278	−0.05961579
	−0.8253307	0.118970201	−0.55197406

	−0.5010919	−0.75922724	−0.3789449
$\boldsymbol{Q}_1$	−0.7655473	0.38089150	0.3981503
	0.1853278	−0.52714993	0.8284170
	−0.3584660	−0.02467012	0.1077147

To check that the representation in (17.2) is accurate we can perform the following product:

svd(**A**)\$u% ∗ %diag(svd(**A**)\$d)% ∗ %t(svd(**A**)\$v)

where \$d is the symbol used by R to denote the singular values, \$u and \$v are the symbols used to denote the matrices $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$, respectively. It can be verified that this product is equal to the matrix $\boldsymbol{A}$.
7. **The Cholesky Decomposition**. This decomposition was described by Theorem 6.3. Thus if $\boldsymbol{A}$ is, for example, positive definite, then there exists a unique upper triangular matrix $\boldsymbol{T}$ with positive diagonal elements such that $\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}$. This decomposition can

be implemented in R by using the function *chol*($\boldsymbol{A}$). To demonstrate this, consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Using *chol*($\boldsymbol{A}$) gives the matrix

$$\boldsymbol{T} = \begin{bmatrix} 1.414214 & 1.414214 & 0.7071068 \\ 0 & 1.732051 & 0 \\ 0 & 0 & 1.224745 \end{bmatrix}.$$

It can be verified that $\boldsymbol{A} = \boldsymbol{T}' \boldsymbol{T}$.

8. **Determinant, Inverse, and Generalized Inverse**. The determinant of a square matrix $\boldsymbol{A}$ in R is denoted by *det*($\boldsymbol{A}$). The inverse of $\boldsymbol{A}$ in R is denoted by *solve*($\boldsymbol{A}$). The Moore–Penrose generalized inverse of a matrix $\boldsymbol{B}$, denoted by *ginv*($\boldsymbol{B}$), requires the loading of the MASS package. However, there is no ready-made function to produce just a generalized inverse that is not unique (see, for example, Fieller (2016, page 128).
9. **Rank of a Matrix**. The rank of a matrix $\boldsymbol{A}$ is obtained by first inputting the matrix, then typing *qr*($\boldsymbol{A}$)\$*rank* (see Example 17.8). The rank can also be obtained indirectly by taking the product $\boldsymbol{A}\% * \%t(\boldsymbol{A})$ which is symmetric. The number of nonzero eigenvalues of this product is equal to its rank and hence the rank of $\boldsymbol{A}$ (see Section 6.5.4).
10. **Writing a Comment**. The symbol # is used to write a comment in R. Everything to the end of a line after # is a comment.
11. **Downloading Packages**. All R functions are stored in packages. To perform any task in R requires the availability of an appropriate package that pertains to the task at hand. There are hundreds of contributed packages written by many contributors. These packages perform a variety of statistical procedures and data manipulations. To install packages, the following steps can be followed:
    - open R by double clicking on the R icon.
    - click on “Packages” on top in the tool bar.
    - click on “Install package(s)”.
    - Scroll down to USA (CA2).
    - select the desired program(s), for example, DAAG for data analysis and graphics data.

    You can find out what packages are available by going to the site “cran.stat.ucla.edu” which gives “The Comprehensive R Archive Network.” By clicking on “Packages” on the left and then clicking on “Table of available packages, sorted by name,” you get a list of all packages to be used and what they are used for, for example, DAAG mentioned earlier and lmm for linear mixed models.

    Finding packages can also be done by using the *Task Views* on CRAN (http://cran.r-project.org/web/views). This gives a listing of packages within a particular application area, for example, Bayesian for Bayesian inference. To install and load a particular package, such as RSA for response surface analysis, we use the command

install.packages("RSA")

This is followed by scrolling down to USA (CA2), for example, then typing the command

library(RSA)

We can find out what packages are available on one's computer by starting up R and typing library().

12. **Saving Files**. To save a file, click on File in the upper left of the screen and select "Save to File". Enter the name of the file and press "save" in the chosen folder. To display a file already saved, click on File and select "Display file". Click on the name of the file, then click on "Open".
13. **Storing and Removing Objects**. In a given R session, objects can be created and stored by name. In order to find out what objects are currently stored within R, we use the command "objects()". The same result can be arrived at by using the "ls()" command. To remove objects, we use the function "rm", for example, rm(x,y) removes x and y from the workspace. Objects created during an R session are stored permanently, that is, if x is used in one session and x in another session, R uses the first x value, which can be confusing. So, x should be removed at the end of a session by typing rm(x). When R is started later on, the workspace from the previous session is reloaded. It is therefore advisable to have separate working directories for each analysis carried out using R. The entire workspace can be removed by using rm(list=ls()).
14. **Reading in Numerical Data**. A data set can be entered directly from the keyboard by using the scan function, scan(). For example, we can type x=scan(). If we press the enter key, we get a line where we can enter the data. Hitting the press key again we get another line for data entry, and so on. When we complete entering the data, we can press the enter key twice which will have the effect of terminating the scanning. Typing x on the command line gives us the entire data set that was entered.
15. **Some Mathematical and Statistical Functions**. Some mathematical functions, such as log(x), the natural logarithm of x; log10(x), the common logarithm of x; exp(x), the exponential function; sqrt(x), the square root function; abs(x), the absolute value of x; factorial(x); and the trigonometric functions, sin(x), cos(x), tan(x), acos(x), asin(x), atan(x) are available in R, in addition to min(x) and max(x) for the smallest and the largest elements of a vector. The function length(x) represents the number of elements in x and sum(x) represents the total of its elements. As for statistical functions we can include mean(x) which gives the sample mean of the elements in x, and var(x) which gives the sample variance, that is,

$$var(x) = sum((x - mean(x))^2)/(length(x) - 1).$$

sd(x) gives the sample standard deviation, that is, the square root of var(x). Pearson's correlation coefficient between two random variables, x and y, is $cor(x, y)$.

16. **Generation of Random Variables**
    (a) x = rnorm(n) creates n=number of observations from the standard normal distribution. For example, x = rnorm(6) gives -0.9321849, -0.3305639, -1.3592365, 0.1879771, 1.1131102, -0.8996389.

(b) y = rbinom(m, n, p) generates m=number of observations from the binomial distribution with n = number of trials and p = probability of success on each trial. For example, y = rbinom(5,10,0.3) gives 5, 4, 5, 4, 3.

(c) x = rpois(n,lambda) generates n random variables having the Poisson distribution with parameter $\lambda = lambda$. For example, rpois(6,0.8) gives 0, 3, 1, 0, 0, 1.

(d) x = rgamma(n,shape,scale=1) generates n gamma random variables having specified shape and scale parameters. For example, rgamma(6,5,scale=1) generates six gamma random variables, namely, 4.088706, 2.174163, 3.982304, 10.562917, 4.467228, 1.328104 with shape parameter equal to 5 and a scale parameter equal to 1.

(e) x = runif(n,min=0,max=1) generates n random variables from the uniform distribution on the interval from min=0 to max=1. For example, runif(5,min=0,max=1) generates five uniform random variables, namely, 0.1272399, 0.2971816, 0.3252382, 0.3615584, 0.6610594 on the interval from min=0 to max=1. To choose a different interval, for example, an interval from 4.0 to 6.5, use x = runif(n,min=4.0,max=6.5).

To generate n random integers between, for example, 1 and 10, we use the "sample" function. For example, x = sample(1:10, 5, replace=T) generates five random integers between 1 and 10 where sampling can be done with replacement. For example, x = sample(1:10,5,replace=T) gives the integers 8, 4, 10, 6, 4. If we wish to have the sampling done without replacement, we use x = sample(1:10, 5, replace=F) which gives, for example, 6, 1, 4, 5, 2.

(f) x = rchisq(n,df) gives n random variables having the central chi-squared distribution with df degrees of freedom.

(g) x = rt(n,df) gives n random variables having the central t distribution with df degrees of freedom.

(h) x = rf(n,df1,df2) gives n random variables having the central F distribution with df1, df2 degrees of freedom for the numerator and denominator, respectively.

17. **Setting up an Array**. Consider the following 3×2 array:

$$\text{x} = \text{array}(\text{c}(1{:}3,3{:}1), \text{dim} = \text{c}(3,2)),$$

which gives

$$\begin{matrix} 1 & 3 \\ 2 & 2 \\ 3 & 1. \end{matrix}$$

18. **Numerical Summaries of Data**. Suppose we have the vector $x = c(1 : 6, 16, 89)$. Then, the command summary(x) gives the following information about x:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	2.75	4.50	15.75	8.50	89.00

Example 17.1 *Consider the row vector,* $\boldsymbol{A} = c(12, 20, -5, 20, 51, -4)$, *from which we can obtain the matrix,* $\boldsymbol{B} = matrix(\boldsymbol{A}, 2, 3)$, *which gives*

$$\boldsymbol{B} = \begin{bmatrix} 12 & -5 & 51 \\ 20 & 20 & -4 \end{bmatrix}.$$

Consider also the matrix, $\boldsymbol{C} = matrix(1:9,3,3)$*, which is written as*

$$\boldsymbol{C} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

The product $\boldsymbol{B}\% * \%\boldsymbol{C}$ *is the matrix*

$$\boldsymbol{B}\% * \%\boldsymbol{C} = \begin{bmatrix} 155 & 329 & 503 \\ 48 & 156 & 264 \end{bmatrix}.$$

The Kronecker product of $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *is* $\boldsymbol{B}\% \times \%\boldsymbol{C}$ *which is a matrix of order* 6×9.

Example 17.2 *Consider the matrix* $\boldsymbol{A} = matrix(c(1,-3,5,3,0,2,1,7,-4),3,3, byrow = T)$*, which is written as*

$$\boldsymbol{A} = \begin{bmatrix} 1 & -3 & 5 \\ 3 & 0 & 2 \\ 1 & 7 & -4 \end{bmatrix}.$$

We note that the use of byrow = T means that the matrix can be constructed by entering one row after the other. To construct a matrix with just the diagonal elements of $\boldsymbol{A}$*, we use diag(diag(*$\boldsymbol{A}$*)) which gives*

$$diag(diag(\boldsymbol{A})) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

The trace of $\boldsymbol{A}$ *is obtained by using the command sum(diag(*$\boldsymbol{A}$*)), which gives -3. The inverse matrix of* $\boldsymbol{A}$ *is obtained by using the command solve(*$\boldsymbol{A}$*) which gives the matrix*

$$\boldsymbol{A}^{-1} = \begin{bmatrix} -0.2857143 & 0.4693878 & -0.1224490 \\ 0.2857143 & -0.1836735 & 0.2653061 \\ 0.4285714 & -0.2040816 & 0.1836735 \end{bmatrix}.$$

The determinant of $\boldsymbol{A}$ is $det(\boldsymbol{A}) = 49$.
The singular-value decomposition of $\boldsymbol{A}$ is given by the expression in (17.2), where the singular values are
9.747651 4.188422 1.200178

$$\begin{array}{crrr} & 0.5699301 & -0.4432808 & -0.6918683 \\ \boldsymbol{P}_1 & 0.1397001 & -0.7774719 & 0.6132058 \\ & -0.8097305 & -0.4461386 & -0.3811782 \end{array}$$

$$\begin{array}{crrr} & 0.01839418 & -0.7692241 & 0.6387142 \\ \boldsymbol{Q}_1 & -0.75689047 & -0.4281153 & -0.4937956 \\ & 0.65328284 & -0.4743538 & -0.5900932 \end{array}$$

To check that the representation in (17.2) is accurate we can perform the following product:

$$\text{svd}(\mathbf{A})\$\text{u}\% * \%\text{diag}(\text{svd}(\mathbf{A})\$\text{d})\% * \%\text{t}(\text{svd}(\mathbf{A})\$\text{v})$$

where \$d is the symbol used by R to denote the singular values, \$u and \$v are the symbols used to denote the matrices $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$, respectively. It can be verified that this product is equal to the matrix $\boldsymbol{A}$.

Example 17.3 *Consider the two matrices* $\boldsymbol{A} = matrix(\mathbf{c}(1:8), 2, 4)$, $\boldsymbol{B} = matrix(c(1:8), 2, 4, byrow = T)$. *We thus have*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}.$$

To join **A** *and* **B** *together side by side, we use* $cbind(\boldsymbol{A}, \boldsymbol{B})$ *which gives*

$$cbind(\boldsymbol{A}, \boldsymbol{B}) = \begin{bmatrix} 1 & 3 & 5 & 7 & 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 & 5 & 6 & 7 & 8 \end{bmatrix}.$$

To stack **A** *and* **B** *together on top of each other, we use* $rbind(\boldsymbol{A}, \boldsymbol{B})$ *which gives*

$$rbind(\boldsymbol{A}, \boldsymbol{B}) = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}.$$

To find the submatrix **D** *consisting of rows 1 through 2 and columns 1 through 4 of* **C**, *where* $\boldsymbol{C} = rbind(\boldsymbol{A}, \boldsymbol{B})$, *we use* $\boldsymbol{D} = \boldsymbol{C}[1:2, 1:4]$ *which gives*

$$\boldsymbol{D} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}.$$

To find the submatrix **E** *consisting of rows 2 through 4 of* **C**, *we use* $\boldsymbol{E} = \boldsymbol{C}[2:4,]$ *which gives*

$$\boldsymbol{E} = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}.$$

Let us now find the inverse of $\boldsymbol{AB}'$: *Define* $\boldsymbol{F} = \boldsymbol{A}\% * \%t(\boldsymbol{B})$, *then compute* $solve(\boldsymbol{F})$ *which gives*

$$solve(\boldsymbol{F}) = \begin{bmatrix} 0.875 & -0.7125 \\ -0.375 & 0.3125 \end{bmatrix}.$$

It can be verified that $\boldsymbol{F} * \boldsymbol{G} = \boldsymbol{I}_2$ *where* **G** *is the inverse of* **F**.

Example 17.4 *Consider the matrix* $\boldsymbol{A} = matrix[\boldsymbol{c}(4, 1, 1, 1, 4, 1, 1, 1, 3), 3, 3]$. *Let us go through the process of finding the spectral decomposition of* $\boldsymbol{A}$. *To do so, we find the eigenvalues and eigenvectors of* $\boldsymbol{A}$ *by using the command eigen*($\boldsymbol{A}$) *which gives the eigenvalues 5.732051, 3.0, 2.267949 and the corresponding eigenvectors given by the columns of the matrix*

$$\boldsymbol{B} = \begin{bmatrix} -0.6279630 & 0.7071068 & -0.3250576 \\ -0.6279630 & -0.7071068 & -0.3250576 \\ -0.4597008 & 0 & 0.8880738 \end{bmatrix}.$$

Then, the spectral decomposition of $\boldsymbol{A}$ *is given by* $\boldsymbol{BDB}'$, *where* $\boldsymbol{D}$ *is the diagonal matrix of eigenvalues. It can be verified that this product is equal to* $\boldsymbol{A}$. *In R, this product is written as* $eigen(\boldsymbol{A})\$vectors\%{*}\%diag(eigen(\boldsymbol{A})\$values)\%{*}\%t(eigen(\boldsymbol{A})\$vectors)$.

Example 17.5 *Consider the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 10 & 1 \\ 1 & 1 & 4 \end{bmatrix}.$$

This matrix is symmetric and positive definite since its eigenvalues, namely, 10.835021, 4.265809, 2.899170 are all positive. We can therefore apply Cholesky decomposition (see Theorem 6.3) to express $\boldsymbol{A}$ *as*

$$\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T},$$

where

$$\boldsymbol{T} = \begin{bmatrix} 2 & 1 & 0.500 \\ 0 & 3 & 0.1666667 \\ 0 & 0 & 1.9293062 \end{bmatrix}.$$

It can be verified that the product $\boldsymbol{T}'\boldsymbol{T}$ *is indeed equal to* $\boldsymbol{A}$.

Example 17.6 *Consider the matrices*

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix}.$$

The direct product (Kronecker product) of $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *is* $\boldsymbol{C}$, *which is the matrix*

$$\boldsymbol{C} = \boldsymbol{A}\% \times \%\boldsymbol{B}.$$

The eigenvalues of $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are 2,0 for* $\boldsymbol{A}$ *and 2.7320508, -0.7320508 for* $\boldsymbol{B}$. *The eigenvalues of their direct product are 5.464102, -1.464102, 0, 0. It can be seen that the eigenvalues of the direct product are the products of the eigenvalues of* $\boldsymbol{A}$ *with those of* $\boldsymbol{B}$. *This was confirmed by Theorem 6.10. This theorem also states that if* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are eigenvectors of two square matrices,* $\boldsymbol{C}$ *and* $\boldsymbol{D}$, *corresponding to the eigenvalues,* λ *and* μ, *respectively, then*

$\boldsymbol{x} \otimes \boldsymbol{y}$ is an eigenvector of $\boldsymbol{C} \otimes \boldsymbol{D}$ corresponding to $\lambda\mu$. It should be noted, however, that $\boldsymbol{C} \otimes \boldsymbol{D}$ can have eigenvectors that are not written as the direct product of an eigenvector of $\boldsymbol{C}$ and an eigenvector of $\boldsymbol{D}$, that is, we should not always expect that all the eigenvectors of $\boldsymbol{C} \otimes \boldsymbol{D}$ can be reproduced by simply taking all the direct products of the eigenvectors of $\boldsymbol{C}$ with those of $\boldsymbol{D}$. To demonstrate this, let us consider again the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in our present example. The eigenvectors of $\boldsymbol{A}$ are $\boldsymbol{x}_1 = (1, 0)'$ and $\boldsymbol{x}_2 = (-0.7071068, 0.7071068)'$ which correspond to the eigenvalues 2 and 0, respectively. In addition, the eigenvectors of $\boldsymbol{B}$ are $\boldsymbol{y}_1 = (0.9390708, 0.3437238)'$ and $\boldsymbol{y}_2 = (-0.5906905, 0.8068982)'$ which correspond to the eigenvalues 2.7320508 and -0.7320508, respectively. Using R, the eigenvectors of $\boldsymbol{A} \otimes \boldsymbol{B}$, which correspond to the eigenvalues 5.464102, -1.464102, 0, 0, are,respectively, the columns (from left to right) of the following matrix which we denote by $\boldsymbol{V}_c$:

$$\boldsymbol{V}_c = \begin{bmatrix} 0.9390708 & -0.5906905 & -0.7071068 & 0.0000000 \\ 0.3437238 & 0.8068982 & 0.0000000 & -0.7071068 \\ 0.0000000 & 0.0000000 & 0.7071068 & 0.0000000 \\ 0.0000000 & 0.0000000 & 0.0000000 & 0.7071068 \end{bmatrix}.$$

On the other hand, taking the direct product of $\boldsymbol{V}_a$ and $\boldsymbol{V}_b$, where $\boldsymbol{V}_a$ is the matrix whose columns are $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the eigenvectors of $\boldsymbol{A}$ corresponding to 2 and 0, and $\boldsymbol{V}_b$ is the matrix whose columns are $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, the eigenvectors of $\boldsymbol{B}$ corresponding to 2.7320508, -0.7320508, then the direct product $\boldsymbol{V}_a \otimes \boldsymbol{V}_b$ gives the matrix,

$$\boldsymbol{V}_a \otimes \boldsymbol{V}_b = \begin{bmatrix} 0.9390708 & -0.5906905 & -0.6640233 & 0.4176813 \\ 0.3437238 & 0.8068982 & -0.2430494 & -0.5705632 \\ 0.0000000 & 0.0000000 & 0.6640233 & -0.4176813 \\ 0.0000000 & 0.0000000 & 0.2430494 & 0.5705632 \end{bmatrix}.$$

It can be seen that the last two columns in $\boldsymbol{V}_a \otimes \boldsymbol{V}_b$ are eigenvectors of $\boldsymbol{A} \otimes \boldsymbol{B}$ corresponding to the eigenvalue 0. However, from the matrix for $\boldsymbol{V}_c$, its last two columns are also eigenvectors of $\boldsymbol{A} \otimes \boldsymbol{B}$ corresponding to the eigenvalue 0.

The R statements used in this example are

```
A=matrix(c(2,0,2,0),2,2)
B=matrix(c(2,1,2,0),2,2)
C=A%x%B
eigen(A)
Va=eigen(A)$vectors
eigen(B)
Vb=eigen(B)$vectors
eigen(C)
Vc=eigen(C)$vectors
Va%x%Vb
```

*Note that Va, Vb, and Vc are the same as $\boldsymbol{V}_a$, $\boldsymbol{V}_b$, and $\boldsymbol{V}_c$, respectively. The statements, eigen(**A**), eigen(**B**), and eigen(**C**), provide the eigenvalues as well as corresponding eigenvectors for **A**, **B**, and **C**.*

Example 17.7 *Consider constructing the following matrices in R:*

(a) *The column vector of six ones: matrix(rep(1,6),ncol=1), where rep(1,6) means that the value 1 is repeated six times.*
(b) *The identity matrix $\boldsymbol{I}_6$: diag(rep(1,6)).*
(c) *The matrix of ones of order 5×5, $\boldsymbol{J}_5$: matrix(rep(1,25),5,5).*
(d) *The centering matrix of order 4×4, $\boldsymbol{I}_4 - \frac{1}{4}\boldsymbol{J}_4$: diag(rep(1,4)) - matrix(rep(1,16),4,4)/4.*

Example 17.8 *To show that the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 5 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix}$$

is of rank 2.

The rank of $\boldsymbol{A}$ can be determined by using qr(A)\$rank, which gives the value 2 (see Example 17.12). Also, as was stated earlier, to find the rank of $\boldsymbol{A}$ we can take the product $\boldsymbol{A}\% * \%t(\boldsymbol{A})$, which is symmetric, then compute its eigenvalues. The number of nonzero eigenvalues of this product is equal to the rank of $\boldsymbol{AA}'$, and hence the rank of $\boldsymbol{A}$ (see Section 6.5.4). It can be shown that the eigenvalues of $A\% * \%t(A)$ are 1086.2699789,0.7300211,0. Hence, the rank of $\boldsymbol{A}$ must be equal to 2.

Example 17.9 *Given the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 6 \end{bmatrix}.$$

To find the square root of $\boldsymbol{A}$.

Since $\boldsymbol{A}$ is symmetric we can apply its spectral decomposition, which we denote by $\boldsymbol{B}$. We get

$$\boldsymbol{B} = \boldsymbol{P}\% * \%\boldsymbol{\Lambda}\% * \%t(\boldsymbol{P}),$$

where $\boldsymbol{P}$ is the orthogonal matrix of eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. Let $\boldsymbol{S}$ be a diagonal matrix whose diagonal elements are the square roots of the eigenvalues of $\boldsymbol{A}$. These eigenvalues are all positive, namely, 3,4,7. Hence, $\boldsymbol{A}$ is positive definite. The square root, $\boldsymbol{C}$, of $\boldsymbol{A}$ is therefore given by

$$\boldsymbol{C} = \boldsymbol{P}\% * \%\boldsymbol{S}\% * \%t(\boldsymbol{P}),$$

which is the matrix

$$\boldsymbol{C} = \begin{bmatrix} 1.9736508 & 0.2415997 & 0.2152503 \\ 0.2415997 & 1.9736508 & 0.2152503 \\ 0.2152503 & 0.2152503 & 2.4305008 \end{bmatrix}.$$

It can be verified that $\boldsymbol{C}\% * \%\boldsymbol{C}$ is equal to $\boldsymbol{A}$.

Example 17.10 *Consider the matrices,* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, given by*

$$\boldsymbol{A} = \begin{bmatrix} 3 & 8 & 4 & 1 \\ 2 & 4 & -9 & -5 \\ 5 & 3 & 7 & 12 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 12 & 3 & -12 & 4 \\ -4 & 6 & 14 & -9 \\ -11 & 3 & 4 & -6 \end{bmatrix}.$$

To demonstrate that $rank(\boldsymbol{A}+\boldsymbol{B}) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$*:*

As was seen before, to find the rank of $\boldsymbol{A}$ *we compute the nonzero eigenvalues of* $A\% * \%t(A)$*, which are equal to 311.39665, 103.08041, 28.52295. The rank of* $A\% * \%t(A)$ *is equal to 3, which is equal to the rank of* $\boldsymbol{A}$*. Similarly, the nonzero eigenvalues of* $B\% * \%t(B)$ *are 688.07291, 94.24642, 41.68067 and the corresponding rank is 3, which is the same as the rank of* $\boldsymbol{B}$*. Finally, the nonzero eigenvalues of* $(A+B)\% * \%t(A+B)$ *are 475.9051, 320.2583, 192.8366 and the corresponding rank is 3, which is also the rank of* $\boldsymbol{A}+\boldsymbol{B}$*. We conclude that* $rank(\boldsymbol{A}+\boldsymbol{B}) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$*.*

The R statements used in this example are

```
A=matrix(c(3,2,5,8,4,3,4,-9,7,1,-5,12),3,4)
B=matrix(c(12,-4,-11,3,6,3,-12,14,4,4,-9,-6),3,4)
tA=t(A)
tB=t(B)
APA=A%*%t(A)
BPB=B%*%t(B)
ABPAB=(A + B)%*%t(A + B)
eigen(APA)
eigen(BPB)
eigen(ABPAB)
```

Example 17.11 *Consider the following three matrices:*

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 5 & 8 \\ -1 & 2 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 5 & 8 & 9 \\ -3 & 10 & 11 \\ 17 & -6 & 14 \end{bmatrix}.$$

To demonstrate that

$$(\boldsymbol{A}+\boldsymbol{B})\%x\%\boldsymbol{C} = \boldsymbol{A}\%x\%\boldsymbol{C} + \boldsymbol{B}\%x\%\boldsymbol{C}.$$

The R statements needed for this example are

```
A=matrix(c(2,4,3,5),2,2)
B=matrix(c(5,-1,8,2),2,2)
C=matrix(c(5,-3,17,8,10,-6,9,11,14),3,3)
AkC=A%x%C
BkC=B%x%C
ABkC=(A + B)%×%C
D=ABkC - AkC - BkC
```

D=
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0

The matrix $\boldsymbol{D}$ *indicates that* $(\boldsymbol{A}+\boldsymbol{B})\%x\%\boldsymbol{C} = \boldsymbol{A}\%x\%\boldsymbol{C} + \boldsymbol{B}\%x\%\boldsymbol{C}$.

Example 17.12 *Let us now consider the QR decomposition of a full column rank matrix as described in Theorem 5.2. Let* $\boldsymbol{A}$ *be the matrix*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & -1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix}.$$

The appropriate function in R that can produce the QR decomposition of $\boldsymbol{A}$ *is qr(*$\boldsymbol{A}$*). The needed R statements after inputting the matrix are*

*qra = qr(**A**)*
***Q** = qr.Q(qra)*
***R** = qr.R(qra)*
Q
R

The qr.Q(qra) and qr.R(qra) statements are needed to obtain $\boldsymbol{Q}$ *and* $\boldsymbol{R}$. *Typing* $\boldsymbol{Q}$ *we get*

$$\boldsymbol{Q} = \begin{bmatrix} -0.7071068 & -0.5144958 & -0.48505713 \\ 0.0000000 & -0.6859943 & 0.7276069 \\ -0.7071068 & 0.5144958 & 0.4850713 \end{bmatrix}.$$

Typing next $\boldsymbol{R}$ *gives the matrix*

$$\boldsymbol{R} = \begin{bmatrix} -1.414214 & -2.121320 & -0.7071068 \\ 0.000000 & -2.915476 & 0.8574929 \\ 0.000000 & 0.000000 & 2.1828206 \end{bmatrix}.$$

It can be verified that the product Q% ∗ %R gives the matrix $\boldsymbol{A}$. *The qr function also returns the rank of* $\boldsymbol{A}$ *by typing qr(A)\$rank. Thus the QR decomposition is useful in finding the rank of a matrix.*

17.1 TWO- AND THREE-DIMENSIONAL PLOTS

R has very efficient graphics capability since it enables the user to control how the plot should appear. In this section, just like in Chapter 16, we discuss two- and three-dimensional plots.

17.1.1 Two-Dimensional Plots

Example 17.13 *In this example we present a simple plot that shows the fit of a data set with a straight line. The data give the yield, y, of a certain chemical compound which is influenced by the temperature, x, of the reaction. The following R statements are used:*

```
x=c(12,35,50,30,52,66,12,21,31,50,26,42,50)
y=c(120,155,201,175,190,170,140,148,166,180,159,162,195)
plot(x,y,xlab="temperature",ylab="yield")
abline(lm(y ~ x))
```

Note that the last statement is needed to draw the fitted line on the plot. The plot is shown in Figure 17.1.

Example 17.14 *Consider plotting the following ellipse:* $x^2 + 2y^2 = 1$. *The corresponding R statements are*

```
x=seq(-1,1,by=0.01)
plot(x,-(sqrt(0.5)) * sqrt(1 - x^2),type="l",ylim=c(-1,1),ylab="y")
lines(x,(sqrt(0.5)) * sqrt(1 - x^2),type="l")
```

The last statement is for drawing the second half of the ellipse. The plot is shown in Figure 17.2.

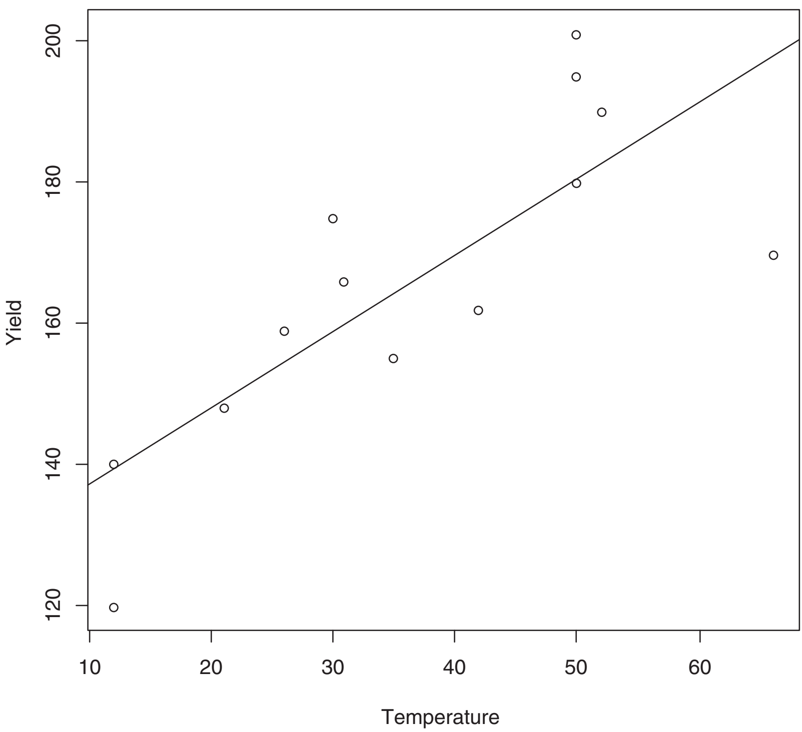


Figure 17.1 *The Yield Data Fitted to a Straight Line.*

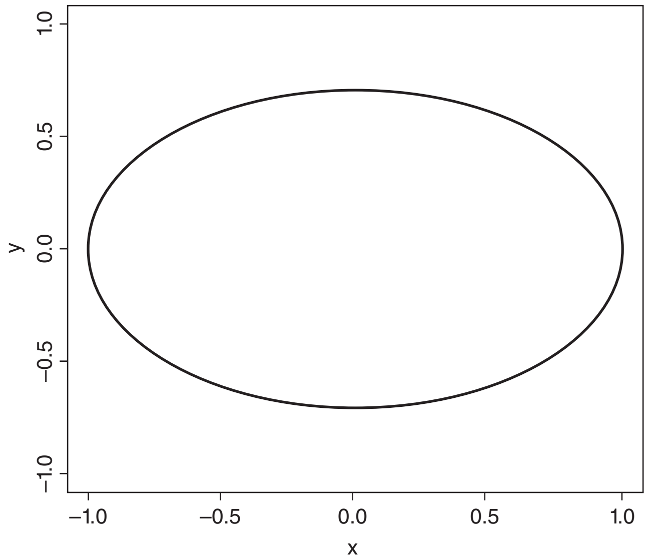


Figure 17.2 *Plot of the Ellipse* $x^2 + 2y^2 = 1$.

Example 17.15 *An experiment was conducted to determine the effects of storage conditions on the quality of a certain variety of apples. Apples were harvested from two different orchards. The quality of apples was measured by the response variable* y = *amount of extractable juice (mL/100 g). The objective here is to compare the two orchards by using their boxplots. The data are shown below:*

Orchard 1	Orchard 2
74.2	73.8
73.1	75.8
74.2	73.6
72.5	71.2
75.0	72.6
74.4	72.9
72.1	70.3
73.3	71.9
74.2	72.0
71.2	70.5
71.9	69.5
72.6	70.5
74.2	72.9
72.8	69.1
70.2	67.7
66.9	64.6

The R statements used are

Orchard1 = scan()
74.2 73.1 74.2 72.5 75.0 74.4 72.1 73.3
74.2 71.2 71.9 72.6 74.2 72.8 70.2 66.9

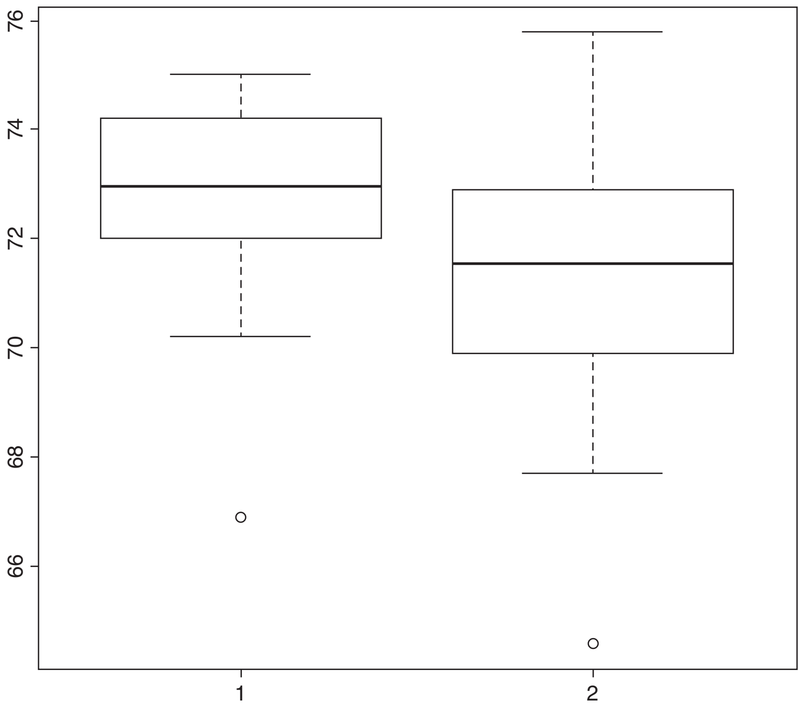


Figure 17.3 *Boxplots for the Two Orchards.*

Orchard2 = scan()
73.8 75.8 73.6 71.2 72.6 72.9 70.3 71.9
72.0 70.5 69.5 70.5 72.9 69.1 67.7 64.6
boxplot(Orchard1,Orchard2)

Note that the data were entered on the keyboard using scan(). The boxplots are shown in Figure 17.3.

The scan function is very convenient to read in numeric data directly from the keyboard. The data can be entered on separate lines. Upon the completion of entering the data, hitting the enter key twice will terminate the scanning.

Example 17.16 *In this example we show how to plot two functions in the same figure. The first function is* $y = cos(x_1) + sin(3\,x_1)$ *with* x_1 *ranging from 0 to 10, and the second function is* $y = 4\,exp(-x_2)\,sin(7\,x_2)\,cos(3\,x_2)$ *with* x_2 *ranging from 0.3 to 10. The R statements used are*

x1=seq(0,10,len=300)
*y1=cos(x1) + sin(3*x1)*
x2=seq(0.3,10,len=300)
*y2=4*exp(-x2)*sin(7*x2)*cos(3*x2)*
*par(mar=**c**(5,4,4,5))*

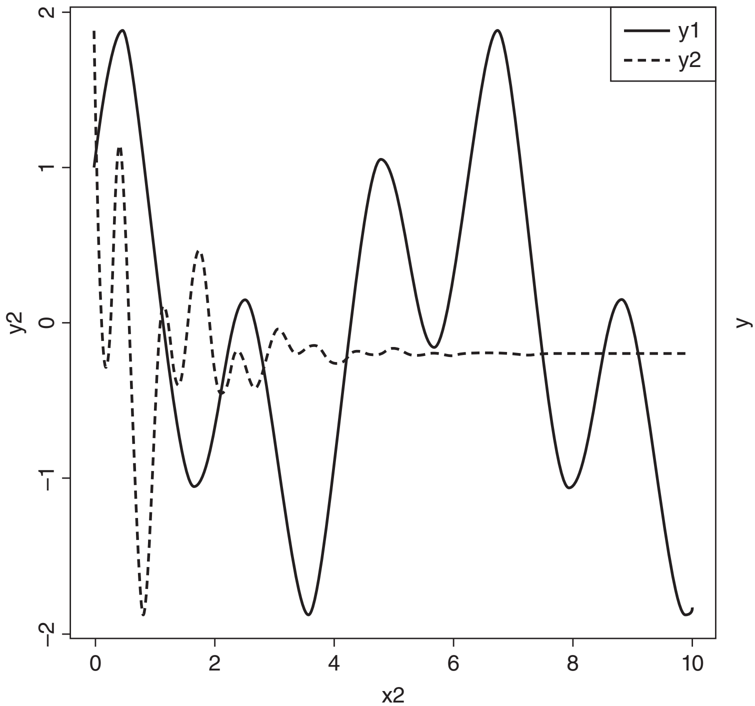


Figure 17.4 *Plot of the Two Functions.*

plot(x1,y1,type='l')
par(new=TRUE)
plot(x2,y2,type='l',lty=2, xaxt='n',yaxt='n')
mtext('y',side=4,line=3)
legend('topright',legend=c('y1','y2'),lty=c(1,2))

The plot of the two functions is shown in Figure 17.4. Note that "lty=2" specifies the line type. The first "par" function, which is used to customize features of the graph, sets the bottom, left, top, and right margins, respectively, of the plot region in number of lines of text. The second "par" function is used to tell R to make the second plot without clearing the first. Furthermore, the "mtext" function places the text in one of the four margins with "side =1, 2, 3, 4 for bottom, left, top, right, respectively." Note that in the statement "par(mar=c(5,4,4,5))", the integers 5, 4, 4, 5 give the number of lines of margin on four sides of the plot (bottom, left, top, right). The statement "par(new=TRUE)" is needed for combining plots. This way, the next plotting command will not clear the frame before drawing the plot. In the "plot(x2,y2,...)" statement, "lty=2" denotes using a dashed line, "xaxt='n'" specifies the x-axis type and using "'n'" suppresses plotting of the axis. A similar explanation applies to "yaxt='n'". Finally, in the "legend" statement, "lty=1" denotes using a solid line.

Example 17.17 *This example shows a plot of a histogram with a normal density. The R statements used are*

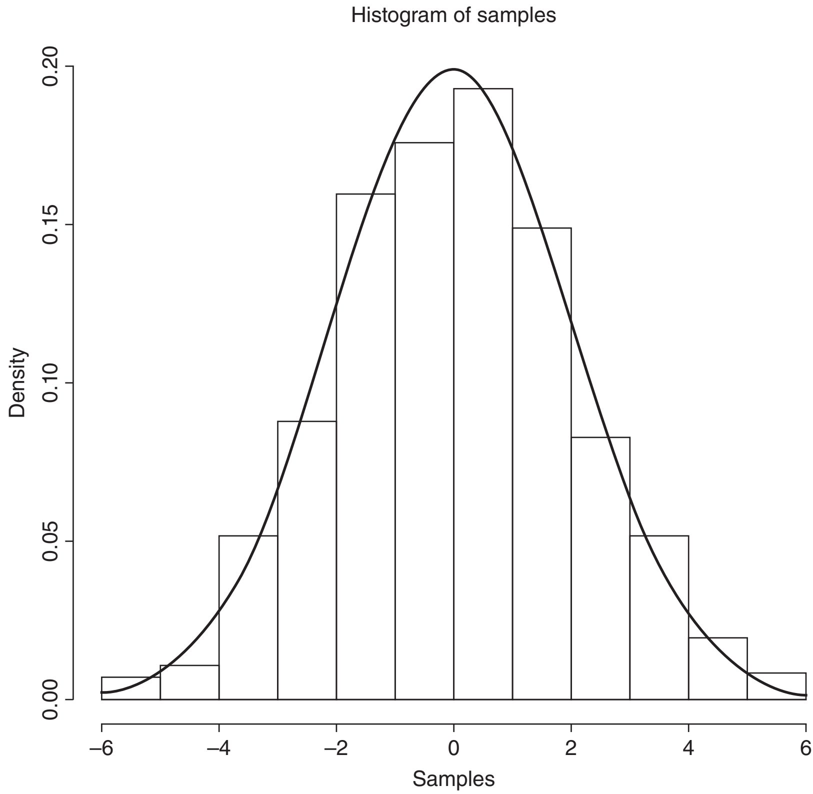


Figure 17.5 *Histogram with the Normal Density Curve.*

```
n=1000
samples=rnorm(n,mean=0,sd=2)
hist(samples,freq=F)
curve(dnorm(x,mean=0,sd=2),add=TRUE)
```

The plot of the histogram is shown in Figure 17.5.

Note that the use of "freq=F" is needed in order to make the histogram represented in terms of relative frequencies, rather than absolute counts, so that the histogram has a total area of one. The curve function gives a representation of an expression expressed in terms of x and the use of "add=TRUE" makes it possible to superimpose a plot over an existing plot.

Example 17.18 *An experiment was conducted to study the effects of two fertilizers, fertilizer A and fertilizer B, on the yield of tomato plants (kg/plot). The following data were collected for three levels of A (20,25,30 kg/hectare), and three levels of B (25,35,45 kg/hectare) and are shown in Table 17.1.*

The R statements used are

```
A=c(20,20,20,20,20,20,25,25,25,25,25,25,30,30,30,30,30,30)
```

TABLE 17.1 Levels of Fertilizers and Corresponding Yield Values

Fertilizer A	Fertilizer B	Yield (kg/plot)
20	25	60
20	25	65
20	35	52
20	35	55
20	45	42
20	45	40
25	25	45
25	25	50
25	35	48
25	35	54
25	45	40
25	45	48
30	25	70
30	25	68
30	35	62
30	35	60
30	45	50
30	45	52

```
B=c(25,25,35,35,45,45,25,25,35,35,45,45,25,25,35,35,45,45)
yield=c(60,65,52,55,42,40,45,50,48,54,40,48,70,68,62,60,50,52)
interaction.plot(A,B,yield)
```

The interaction plot is shown in Figure 17.6.

Example 17.19 *A plot of the correlation matrix is addressed in this example. The plot depicts correlations between variables in a data set. The following R statements concerning three random variables are utilized to do the plotting:*

```
x = sample(1:30,30) + rnorm(15,sd=2)
y = -x + rnorm(15,sd=4)
z = (sample(1:30,30)/3) + rnorm(15,sd=4)
dframe = data.frame(x,y,z)
plot(dframe[,1:3])
```

The values of x, y, and z are generated as random integers between 1 and 30 inclusive from the discrete uniform distribution with added normally distributed errors. A data.frame is a matrix whose columns display different types of data, in this example, the values of x, y, and z. The plot is shown in Figure 17.7.

It is easy to see that x and y are negatively correlated, as they should be, but x and z, and y and z do not to appear to be correlated.

Example 17.20 *This example shows how to draw quantile-quantile plots(q-q plots). These plots are used to graphically assess if a sample of data originates from a normal population. This is accomplished by comparing the quantiles of the sample data to their expected values which are determined by assuming that the sample comes from a standard normal*

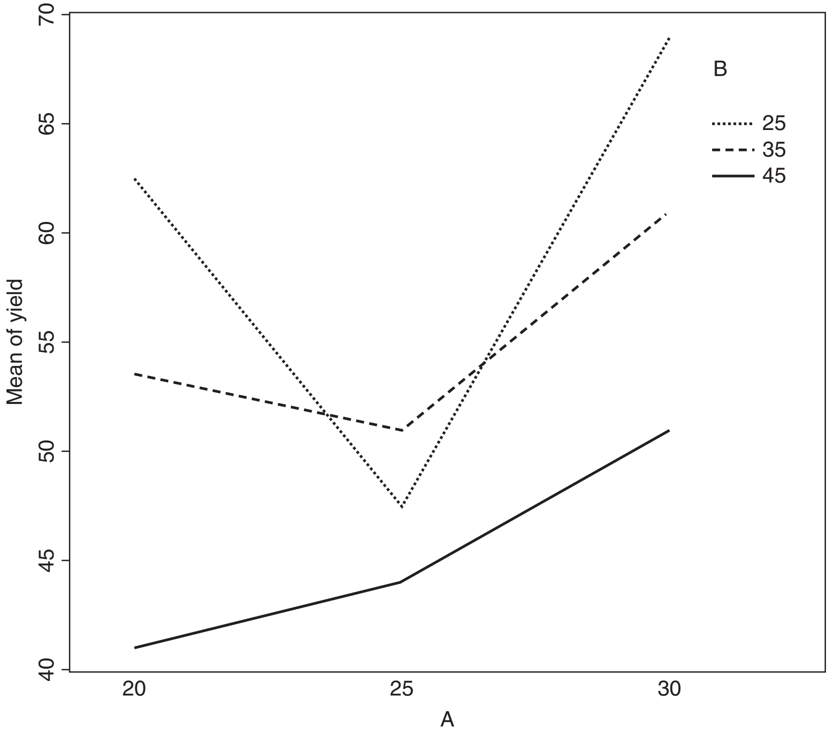


Figure 17.6 *The Fertilizers' Interaction Plot.*

population. The resulting pairs of points produce the so-called q-q plot. If the sample data are normally distributed, the sample quantiles and their corresponding expected values should be reasonably equal. Hence, if the plot is close to being a straight line, then this leads us to believe that the sample possibly comes from a normal population. If, however, the plot is nonlinear, then this gives an indication of a possible problem with the normality assumption regarding the data. This can also be caused by the presence of outliers in the data. A more detailed account of the q-q plots can be found in, for example, Christensen (1996, Section 2.4), Hocking (1996, Section 5.3.3), and Ekstrom (2012, Section 4.17).

It is quite convenient in R to draw the q-q plots using the qqnorm function. There is also the qqline function which adds a comparison line that goes through the first and third quantiles. Departure from this line indicates a problem with the normality assumption.

Let us now consider , for example, two cases: one when the data set comes from a normal distribution, and another when the data come from a lognormal distribution. The relevant R statements are

```
x=rnorm(500,mean=6,sd=3)
y=rlnorm(500,meanlog=6,sdlog=1.1)
qqnorm(x)
qqline(x)
qqnorm(y)
qqline(y)
```

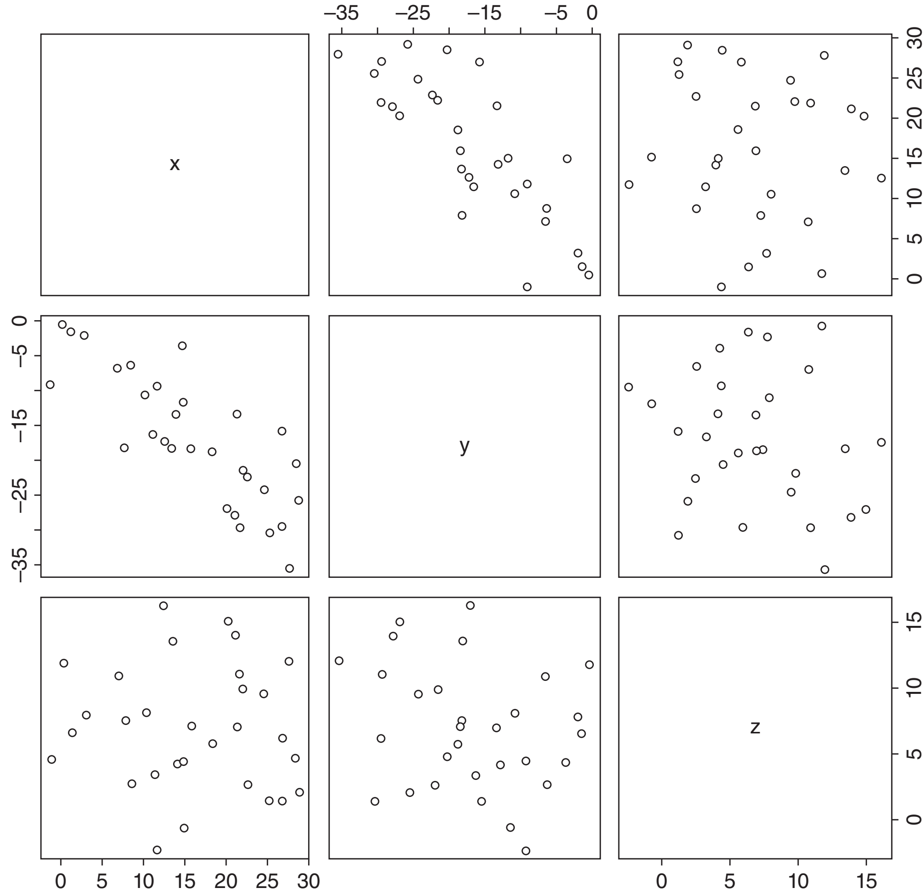


Figure 17.7 *A Plot of a Correlation Matrix.*

The number 500 represents the size of the sample data. The plots for qqnorm(x) and qqnorm(y) are shown in Figures 17.8 and 17.9, respectively.

It can be seen that the plot for the normal distribution follows a straight line, as it should be, while the plot for the lpognormal does not exhibit such a behavior indicating lack of normality.

17.1.2 Three-Dimensional Plots

R has the capability to plot a function $z = f(x, y)$ of two variables, x, y. The function "persp" is used to produce three-dimensional plots where the values of x *and* y can be specified. "ticktype=detailed" is used to specify numerical values on the x, y axes. Furthermore, the parameters "theta" and "phi" are used to specify the viewing angle of the surface.

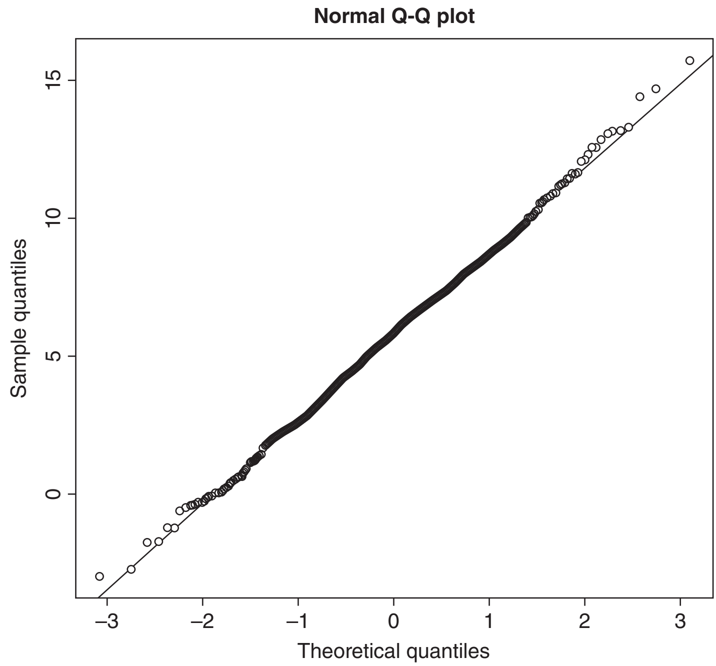


Figure 17.8 *Q-Q Plot for the Normal Distribution.*

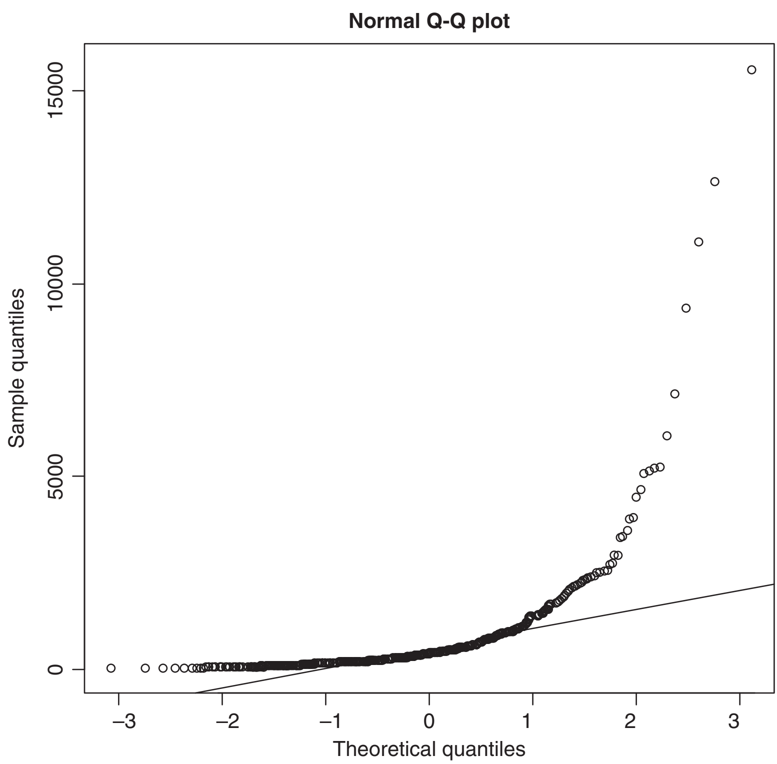


Figure 17.9 *Q-Q Plot for the Lognormal Distribution.*

Example 17.21 *In this example, we show how to plot a quadratic response surface over a bounded region. Consider the model*

$$z = 23.5 + 2x - 3y + 4x^2 + 6y^2 - 8xy.$$

The following R statements are used for the plotting of this function:

```
x=seq(-1,1,0.1)
y=seq(-1,1,0.1)
z=outer(x,y,FUN=function(x,y) (23.5 + 2 * x - 3 * y + 4 * x^2 + 6 * y^2 - 8 * x * y))
persp(x,y,z,theta=30,phi=30,ticktype="detailed")
```

The plot of the function is shown in Figure 17.10.

It would be of interest here to create contour plots of this quadratic surface. For this purpose, the following R statements can be used:

```
x=seq(-1,1,0.1)
y=seq(-1,1,0.1)
z=outer(x,y,FUN=function(x,y) (23.5 + 2 * x - 3 * y + 4 * x^2 + 6 * y^2 - 8 * x * y))
contour(x,y,z,nlevels=10)
```

The "nlevels" option is used to specify the number of contour levels. The contour plots are shown in Figure 17.11.

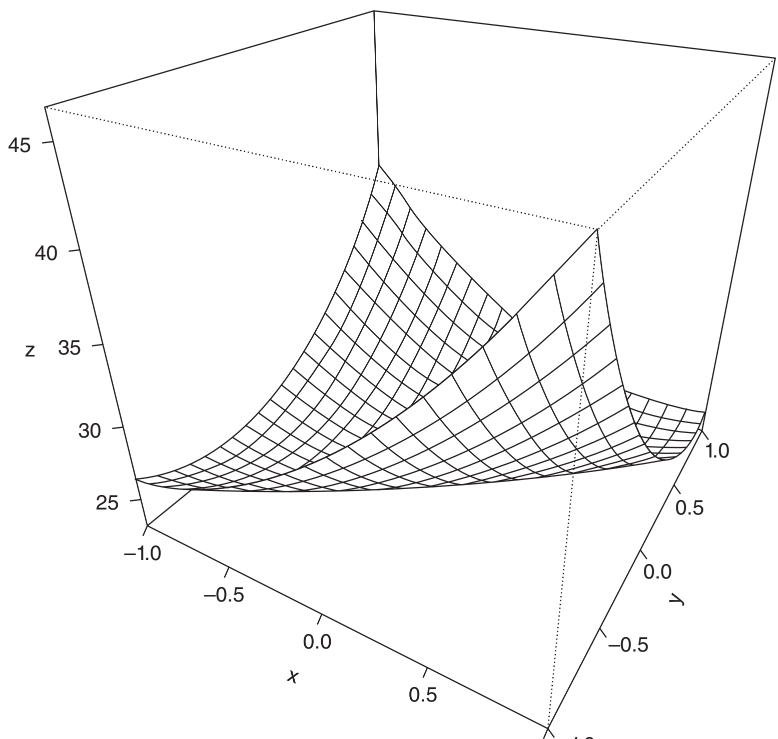


Figure 17.10 *A Quadratic Response Surface.*

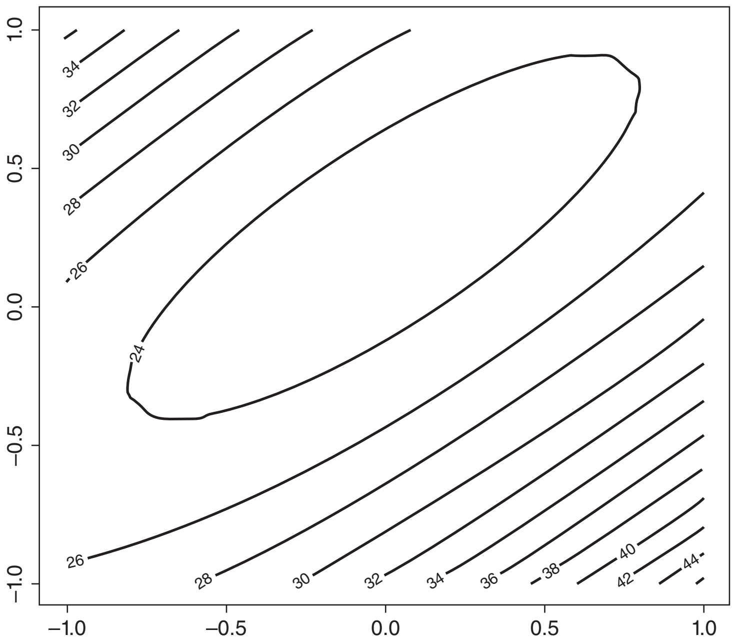


Figure 17.11 *Contour Plots of a Response Surface.*

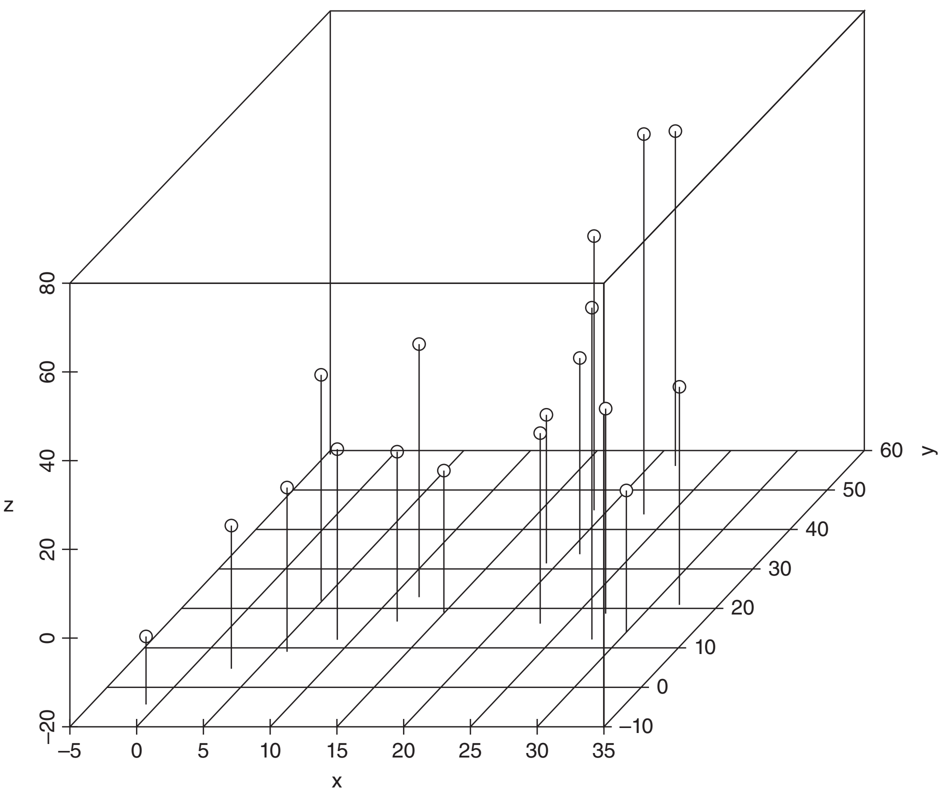


Figure 17.12 *A Three-Dimensional Scatter Plot.*

Example 17.22 *In this example we show how to plot a scatter plot using the function "scatterplot3d". Consider the following R statements:*

```
x=c(-1,3,6,9,12,15,19,23,27,32,5,12,19,21,22,23,28,30)
y=c(-4,4.8,9,12,17,19,32,16,19,21,22,23,45,34,56,44,12,14)
z=c(-5,12,17,23,18,12,13,23,26,29,31,37,42,24,56,66,55,12)
scatterplot3d(x,y,z,type="h")
```

The type="h" option is needed for drawing vertical lines. The plot is shown in Figure 17.12.

There are several useful references for additional reading regarding R. Some of these references include Dalgaard (2002), Venables and Smith (2004), Faraway (2005), Maindonald and Braun (2007), Kleinman and Horton (2010), Vinod (2011), Ekstrom (2012), Verzani (2014), Hiebeler (2015), Unwin (2015), and Fieller (2016). The references by Vinod and Fieller focus on the use of R for doing matrix calculations.

REFERENCES

Christensen, R. (1996). *Analysis of Variance, Design, and Regression*. Chapman & Hall/CRC, Boca Raton.

Dalgaard, P. (2002). *Introductory Statistics with R*. Springer, New York.

Ekstrom, C. T. (2012). *The R Primer*. Chapman & Hall/CRC, Boca Raton.

Faraway, J. J. (2005). *Linear Models with R*. Chapman & Hall/CRC, Boca Raton.

Fieller, N. (2016). *Basics of Matrix Algebra for Statistics with R*. Chapman & Hall/CRC, Boca Raton.

Hiebeler, D. E. (2015). *R and MATLAB*. Chapman & Hall/CRC, Boca Raton.

Hocking, R. R. (1996). *Methods and Applications of Linear Models*. John Wiley & Sons, New York.

Kleinman, K. and Horton, N. J. (2010). *SAS and R*. Chapman & Hall/CRC, Boca Raton.

Maindonald, J. and Braun, J. (2007). *Data Analysis and Graphics Using R*, 2nd ed. Cambridge University Press, New York.

SAS Institute, Inc. (2004). *SAS/IML 9.1 User's Guide*. SAS Institute, Inc., Cary, NC.

Unwin, A. (2015). *Graphical Data Analysis with R*. Chapman & Hall/CRC, Boca Raton.

Venables, W. N. and Smith, D. M. (2004). *An Introduction to R*. Network Theory Ltd., Bristol, UK.

Verzani, J. (2014). *Using R for Introductory Statistics*. Chapman & Hall/CRC, Boca Raton.

Vinod, H. D. (2011). *Hands-On Matrix Algebra Using R*. World Scientific, Singapore.

EXERCISES

Do the following exercises using R. Some of these exercises can be compared with those in Chapter 16.

17.1 Consider the matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \begin{bmatrix} 8 & 1 & 3 \\ 10 & 7 & 11 \\ 4 & 5 & 10 \end{bmatrix}.$$

(a) Find the rank of $\boldsymbol{A}$.

(b) Find the determinant of $\boldsymbol{A}$.

(c) Find the eigenvalues of $\boldsymbol{A}$.

(d) Find the trace of $\boldsymbol{A}$.

17.2 Let $\boldsymbol{A}$ be the matrix

$$\boldsymbol{A} = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 5 \end{bmatrix}$$

(a) Find the eigenvalues of $\boldsymbol{A}$ and conclude that it is positive definite.

(b) Find the inverse of $\boldsymbol{A}$.

(c) Find the trace of $\boldsymbol{A}$ in two ways.

17.3 Consider again the matrix in Exercise 2.

(a) Verify that its leading principal minors are all positive.

(b) Verify that $\boldsymbol{A}'\boldsymbol{A}$ is positive definite.

17.4 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$$

(a) Verify that $\boldsymbol{A}$ is positive semidefinite.

(b) Give values of the leading principal minors.

(c) Give values of two principal minors of order 2.

17.5 Consider once more the matrix $\boldsymbol{A}$ in Exercise 4.

(a) Find the trace of $\boldsymbol{A}^2$.

(b) Verify that $tr(\boldsymbol{A}^2) \leq [tr(\boldsymbol{A})]^2$.

17.6 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 4 & -2 \\ -4 & 0 & -8 \\ 2 & 8 & 0 \end{bmatrix}$$

(a) Show that $\boldsymbol{A}$ is skew-symmetric.

(b) Verify that $(\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} + \boldsymbol{A})^{-1}$ is orthogonal.

17.7 Consider the positive definite matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

(a) Find the Cholesky decomposition of $\boldsymbol{A}$.

(b) Apply the spectral decomposition theorem to $\boldsymbol{A}$.

17.8 Consider the matrix

$$A = \begin{bmatrix} 9 & -4 \\ 3 & -12 \\ 6 & -9 \end{bmatrix}.$$

(a) Find the eigenvalues of $A'A$.

(b) Verify that $det(C) \leq \prod_{i=1}^{2} c_{ii}$, where c_{ii} is the ith diagonal element of $C = A'A$.

(c) Find the singular-value decomposition of A.

17.9 Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 7 \\ 3 & -1 & 9 \\ 5 & -4 & 1 \end{bmatrix}.$$

(a) Find the rank of A.

(b) Find the QR decomposition of A.

(c) Verify that the decomposition in part (b) is correct.

17.10 Consider the following matrices:

$$A = \begin{bmatrix} 3 & 5 & 8 \\ 6 & 1 & 2 \\ 9 & 4 & 3 \\ 2 & 0 & 10 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 & -3 \\ -5 & 6 & 2 \\ 2 & 10 & 4 \\ 8 & -5 & 2 \end{bmatrix}.$$

Verify that

$$[tr(A'B)]^2 \leq tr(A'A)\, tr(B'B),$$

which is the Cauchy–Schwarz inequality for matrices.

17.11 Consider the matrix

$$A = \begin{bmatrix} 10 & 2 & 6 \\ 2 & -11 & -8 \\ 6 & -8 & 12 \end{bmatrix}.$$

(a) Find the Euclidean norm of A (this is the square root of the sum of squares of all the elements of A).

(b) Verify that the square of the Euclidean norm of A is equal to the sum of squares of the eigenvalues of A.

17.12 Consider the two matrices

$$A = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 3 & 4 \end{bmatrix}.$$

(a) Find the direct product of $\boldsymbol{A}$ and $\boldsymbol{B}$.

(b) Verify that $|\boldsymbol{A} \otimes \boldsymbol{B}| = |\boldsymbol{A}|^2 |\boldsymbol{B}|^3$.

17.13 Consider the two matrices

$$\boldsymbol{A} = \begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}.$$

(a) Verify that $tr[(\boldsymbol{AB})^2] \leq tr(\boldsymbol{A}^2\boldsymbol{B}^2)$.

(b) Verify that $tr(\boldsymbol{AB}) \leq \frac{1}{2} tr(\boldsymbol{A}^2 + \boldsymbol{B}^2)$.

17.14 Plot the function,

$$y = sin(x)\, cos(x) + cos(3\, x),$$

over the interval $[0, 6.28]$.

17.15 Plot the function,

$$z = y/(1 + 3\, x^2 + 7\, y^2),$$

over the range [-2,2] for each of x and y.

Appendix

Solutions to Exercises

CHAPTER 1

1.1 **(a)** Let a_1, a_2, a_3 be scalars such that $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + a_3\boldsymbol{x}_3 = \mathbf{0}$. Use the coordinates of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ to derive three linear equations which can be solved for the three scalars. The determinant of the coefficients of the scalars in these equations is equal to zero, hence we can have many nonzero solutions. One of these solutions is $a_1 = -8$, $a_2 = 5$, $a_3 = -1$. Hence, the three vectors are linearly dependent.

(b) Applying a similar methodology to derive three linear equations in the scalars b_1, b_2, b_3 using the coordinates of x_1, x_2, x_4, we conclude that all the scalars must be zero since the determinant of the coefficients of the scalars is equal to -2, which is not zero. Hence, the three vectors must be linearly independent.

The linear combination that equals $(a, b, c)'$ is $b_1 = 2c - b + a$, $b_2 = \frac{1}{2}(b - a - c)$, $b_3 = \frac{1}{2}(-a + 3b - 5c)$.

1.2 The sum of two elements in $U \times V$ must belong to $U \times V$ since U and V are vector spaces. Also, the product of a scalar by an element $(\boldsymbol{u}, \boldsymbol{v})$ in $U \times V$ must belong to $U \times V$. Its zero element is $(0_1, 0_2)$ where 0_1 and 0_2 are the zero elements in U and V, respectively. All other properties of Definition 1.1 are satisfied. Hence, $U \times V$ is a vector space.

1.3 **(a)** Suppose that we have two scalars, a_1 and a_2, such that $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 = \mathbf{0}$. Using the coordinates of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, we get the only solution $a_1 = a_2 = 0$ to the resulting three linear equations. Hence, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are linearly independent.

(b) It is easy to show that each of $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ is a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Hence, V has dimension 2.

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e

1.4 Suppose there exist $b_1, b_2, \dots, b_n$ such that $\boldsymbol{u} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i = \sum_{i=1}^{n} b_i \boldsymbol{u}_i$. Then, $\sum_{i=1}^{n}(a_i - b_i)\boldsymbol{u}_i = \boldsymbol{0}$. This is only possible if $a_i - b_i = 0$ for all $i = 1, 2, \dots, n$ since the $\boldsymbol{u}_i's$ are linearly independent.

1.5 If $U \subset V$ and $U \neq V$, then there exists an element $\boldsymbol{w} \in V$, but not in U. If $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_m$ is a basis for U, then the linear span consisting of $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_m, \boldsymbol{w}$ is a subset of V which contains U. Hence, $\dim(U) < \dim(V)$.

1.6 Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \in L(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$. Then, $\boldsymbol{v}_1 = \sum_{i=1}^{m} a_i \boldsymbol{u}_i$, $\boldsymbol{v}_2 = \sum_{i=1}^{m} b_i \boldsymbol{u}_i$. Hence, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \sum_{i=1}^{m}(a_i + b_i)\boldsymbol{u}_i$ which belongs to $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$. Also, if $\boldsymbol{v} \in L(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$ and α is a scalar, then $\alpha\boldsymbol{v} = \sum_{i=1}^{m}(\alpha\gamma_i)\boldsymbol{u}_i$ belongs to $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$. Thus, $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$ is a vector subspace of U.

1.7 **(a)** If $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \aleph(T)$, then $T(\boldsymbol{u}_1) = \boldsymbol{0}$ and $T(\boldsymbol{u}_2) = \boldsymbol{0}$. Hence, if a and b are scalars, then $T(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) = aT(\boldsymbol{u}_1) + bT(\boldsymbol{u}_2) = \boldsymbol{0}$. We conclude that $a\boldsymbol{u}_1 + b\boldsymbol{u}_2$ belongs to the null space which is therefore a vector subspace of U.

(b) If $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \Re(T)$, then there exist $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$ such that $T(\boldsymbol{u}_1) = \boldsymbol{v}_1$ and $T(\boldsymbol{u}_2) = \boldsymbol{v}_2$. Thus, if c and d are scalars, then $c\boldsymbol{v}_1 + d\boldsymbol{v}_2 = T(c\boldsymbol{u}_1 + d\boldsymbol{u}_2) \in \Re(T)$ since $c\boldsymbol{u}_1 + d\boldsymbol{u}_2 \in U$. Hence, $\Re(T)$ is a vector subspace of V.

1.8 $x_1 + 3x_2 = 0$ and $2x_3 - 7x_4 = 0$ are two linearly independent relationships among the four variables. Hence, the dimension of this vector space is 2.

1.9 The null space consists of all x_1, x_2, x_3 such that $3x_1 - 4x_2 + 9x_3 = 0$. This represents a plane in a three-dimensional space that goes through the origin. Hence, the dimension of the null space is 2.

1.10 Suppose that whenever $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are linearly independent, then the $T(\boldsymbol{u}_i)'s$ are linearly independent in V. To show that T is one-to-one, let $\boldsymbol{u} \in U$ be such that $T(\boldsymbol{u}) = \boldsymbol{0}_2$, the zero element in V. To show that $\boldsymbol{u} = \boldsymbol{0}_1$, the zero element in U. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ be a basis for U. Then, $\boldsymbol{u} = \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i$ for some scalars, $\alpha_1, \alpha_2, \dots, \alpha_n$, not all equal to zero (if the scalars are all zero, then $\boldsymbol{u} = \boldsymbol{0}_1$ and we are done). We now have $\boldsymbol{0}_2 = T(\boldsymbol{u}) = \sum_{i=1}^{n} \alpha_i T(\boldsymbol{e}_i)$. This implies that $\alpha_i = 0$ for all i since the $T(\boldsymbol{e}_i)$'s are linearly independent. It follows that $\boldsymbol{u} = \boldsymbol{0}_1$ and T is therefore one-to-one.

Vice versa, suppose that T is one-to-one. To show that whenever $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are linearly independent, then the $T(\boldsymbol{u}_i)'s$ are linearly independent: If this condition is not satisfied, then there exist scalars, $b_1, b_2, \dots, b_n$, not all equal to zero such that $b_1 T(\boldsymbol{u}_1) + b_2 T(\boldsymbol{u}_2) + \cdots + b_n T(\boldsymbol{u}_n) = \boldsymbol{0}_2$. This implies that $T(b_1\boldsymbol{u}_1 + b_2\boldsymbol{u}_2 + \cdots + b_n\boldsymbol{u}_n) = \boldsymbol{0}_2$. Hence, $b_1\boldsymbol{u}_1 + b_2\boldsymbol{u}_2 + \cdots + b_n\boldsymbol{u}_n = \boldsymbol{0}_1$, which indicates that the $\boldsymbol{u}_i$'s are not linearly independent, a contradiction. The $T(\boldsymbol{u}_i)$'s must therefore be linearly independent.

1.11 Suppose not, then there exist scalars, a and b, such that $ax + be^x = 0$ for all x in $[0, 1]$. For $x = 0$, $b = 0$, which implies that $ax = 0$. Letting now $x = 1$, we get $a = 0$. Hence, x and e^x are linearly independent.

CHAPTER 2

2.1 These are straightforward calculations.

2.2 $$A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 5 & 7 & 9 & 11 \\ 6 & 8 & 10 & 12 & 14 \\ 9 & 11 & 13 & 15 & 17 \\ 12 & 14 & 16 & 18 & 20 \end{bmatrix}.$$

2.3 **(a)** The first sum of squares on the left gives the sum of squares of the elements in each row. This sum of squares is then added for all rows. Obviously, this should be equal to the right-hand side which takes the sum over i first then with respect to j.

(b) True since $\sum_{j=1}^{n} a_{ij} = a_{i.}$.

(c) The left-hand side is equal to $(\sum_{j=1}^{n} a_{ij})(\sum_{k=1}^{n} a_{hk})$, which is equal to $a_{i.}a_{h.}$, for $i \neq h$.

(d) This is straightforward.

2.4 Use the following MATLAB statements (see Chapter 16):
$\boldsymbol{A} = [13\ 4\ 6\ 16;\ 7\ 8\ 11\ 9;\ 9\ 10\ 8\ 15;\ 6\ 12\ 15\ 2]$
$a1 = \boldsymbol{A}(1:4, 1)$
$a2 = \boldsymbol{A}(1:4, 2)$
$a3 = \boldsymbol{A}(1:4, 3)$
$a4 = \boldsymbol{A}(1:4, 4)$
$d = [1; 2; 3; 4]$
plot$(d, a1, '+-', d, a2, 'o--', d, a3, 'k*:', d, a4, 'd-.')$.

2.5 Use the following MATLAB statements (see Chapter 16):
$\boldsymbol{A} = [2\ 5\ 7;\ 0\ 7\ 3;\ 0\ 0\ 9]$
$a1 = \boldsymbol{A}(1:3, 1)$
$a2 = \boldsymbol{A}(1:3, 2)$
$a3 = \boldsymbol{A}(1:3, 3)$
$d = [1; 2; 3]$
plot$(d, a1, '+-', d, a2, 'o--', d, a3, 'd-.')$.

2.6 $$A = \begin{bmatrix} 4 & 9 & 16 \\ 9 & 16 & 25 \\ 16 & 25 & 36 \end{bmatrix}.$$

It is obviously symmetric due to the definition of the a_{ij}'s.

2.7 **(a)** $\boldsymbol{D}_1 = \text{diag}(\frac{1}{3}, 1, 3, 9)$.

(b) $\boldsymbol{D}_2 = \text{diag}(\frac{4}{3}, 3, 6, 13)$.

2.8 The number of elements above the diagonal of $\boldsymbol{A}$ is $\frac{1}{2}n(n-1)$.

2.9 $2a_{ii} = 0$ for all i, that is, the diagonal has zero elements. Hence, $\boldsymbol{A}$ can have, in general, $\frac{1}{2}n(n-1)$ different elements above its diagonal.

2.10 $x = 4$, $y = -\frac{1}{2}$, $z = -1$, $w = \frac{7}{2}$.

CHAPTER 3

3.1 Both parts check out.

3.2 **(a)** The columns of $\boldsymbol{x}\mathbf{1}'$ are identic al, hence, the determinant must be equal to zero.

(b) Let a_{ij} be the (i,j)th element of the $n \times n$ matrix $\boldsymbol{A}$. Then, $\boldsymbol{A}\mathbf{1} = \mathbf{0}$ implies that for each i, $a_{in} = -\sum_{j=1}^{n-1} a_{ij}$, that is, the nth column of $\boldsymbol{A}$ is the sum of the first $n-1$ columns multiplied by a minus. Now, the determinant of $\boldsymbol{A}$ is not changed if the first $n-2$ columns of $\boldsymbol{A}$ are added to the $(n-1)$th column by Theorem 3.3. This makes this column equal to the nth column, except for the sign. Hence, by Theorem 3.2, the determinant of $\boldsymbol{A}$ must be equal to zero.

3.3 **(a)** $|\boldsymbol{A}| = a^2f^2 + 2acdf - 2abef + b^2e^2 - 2bcde + c^2d^2$.

(b) This equality checks out.

3.4 **(a)** Let the left-hand side of this equation be denoted by $|\boldsymbol{A}|$. Then,

$$|\boldsymbol{A}| = -\begin{vmatrix} x & x & x \\ -2 & 1 & 0 \\ 7 & 4 & 5 \end{vmatrix}.$$

Adding now the third row to the second row causes the elements in the second row to be equal to 5. Hence, choosing $x = 5$ makes the determinant equal to zero.

(b) Choosing $x = 2$ makes the first two rows identical and the determinant will be equal to zero.

3.5 The area of the triangle ABC is

$$\begin{aligned}\text{Area of triangle} &= \frac{1}{2}[(y_1 + y_2)(x_2 - x_1) + (y_3 + y_2)(x_3 - x_2) - (y_1 + y_3)(x_3 - x_1)] \\ &= \frac{1}{2}(x_2y_1 - x_1y_2 - x_2y_3 + x_3y_2 - x_3y_1 + x_1y_3).\end{aligned}$$

Computing $\frac{1}{2}|\boldsymbol{M}|$, we get

$$\frac{1}{2}|\boldsymbol{M}| = -\frac{1}{2}(x_3y_2 - x_2y_3 + x_1y_3 - x_1y_2 - x_3y_1 + x_2y_1).$$

Hence, the area of the triangle is the absolute value of $\frac{1}{2}|\boldsymbol{M}|$.

3.6 The equation of a line passing through (x_1, y_1) and (x_2, y_2) is

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1},$$

that is,

$$x_2y - x_1y - x_2y_1 - xy_2 + x_1y_2 + xy_1 = 0.$$

This is the same as the expansion of the determinant equated to zero.

3.7 **(a)** The solution is $x_1 = 1, x_2 = 2, x_3 = 7$.
(b) This is true since $|\boldsymbol{A}| = -6$, $|\boldsymbol{A}1| = -6$, $|\boldsymbol{A}_2| = -12$, $|\boldsymbol{A}_3| = -42$.
(c) $|\boldsymbol{A}| \neq 0$.

3.8
$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & b-a & c-a \\ a^2 & b^2 & c^2 \end{vmatrix},$$
which is equal to $(b-a)(c-a)(c-b)$.

3.9
$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ bc & ac & ab \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & b-a & c-a \\ bc & ac & ab \end{vmatrix},$$
which is equal to $(b-a)(c-a)(c-b)$.

3.10 If $\sum_{i=1}^{n} a_{ij} = 1$ for $i = 1, 2, \ldots, n$, then $\boldsymbol{A}\mathbf{1} = \mathbf{1}$, that is, $(\boldsymbol{A} - \boldsymbol{I})\mathbf{1} = \mathbf{0}$. Applying the result in Exercise 3.2(b), we conclude that $|\boldsymbol{A} - \boldsymbol{I}| = 0$.

3.11
$$\begin{vmatrix} a_1 + \lambda\alpha_1 & b_1 + \lambda\beta_1 & c_1 + \lambda\gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} \lambda\alpha_1 & \lambda\beta_1 & \lambda\gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$
Now apply Theorem 3.5 to the second determinant on the right.

3.12 By applying the hint, the second term on the right is equal to zero since two of its rows will be identical.

3.13 $|\boldsymbol{K}|^5 = 3^{12}|\boldsymbol{K}|$. Either $|\boldsymbol{K}| = 0$ or $|\boldsymbol{K}|^4 = 3^{12}$. Hence, $(|\boldsymbol{K}|^2 - 3^6)(|\boldsymbol{K}|^2 + 3^6) = 0$. The latter equation gives $|\boldsymbol{K}| = \pm 3^3$.

3.14 Consider the equality
$$\begin{bmatrix} \boldsymbol{I}_m - \boldsymbol{AB} & \boldsymbol{A} \\ \mathbf{0} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_m & \mathbf{0} \\ \boldsymbol{B} & \boldsymbol{I}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_m & \mathbf{0} \\ \boldsymbol{B} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{A} \\ \mathbf{0} & \boldsymbol{I}_n - \boldsymbol{BA} \end{bmatrix}.$$
Taking the determinants on both sides, we get
$$|\boldsymbol{I}_m - \boldsymbol{AB}| = |\boldsymbol{I}_n - \boldsymbol{BA}|.$$

3.15 We have
$$|\lambda\boldsymbol{I}_n + \mathbf{1}_n\boldsymbol{a}'| = \lambda^n \left| \boldsymbol{I}_n - \left(-\frac{\mathbf{1}_n}{\lambda}\right)\boldsymbol{a}' \right|.$$
Using the result in Exercise 3.14, with $\boldsymbol{A} = -\mathbf{1}_n/\lambda$, $\boldsymbol{B} = \boldsymbol{a}'$, we get
$$\lambda^n \left| \boldsymbol{I}_n - \left(-\frac{\mathbf{1}_n}{\lambda}\right)\boldsymbol{a}' \right| = \lambda^n |1 - \boldsymbol{a}'(-\mathbf{1}_n/\lambda)| = \lambda^{n-1}(\lambda + \boldsymbol{a}'\mathbf{1}_n).$$

3.16 Consider the equality

$$|a\boldsymbol{I}_n + b\boldsymbol{J}_n| = a^n |\boldsymbol{I}_n + (b/a)\boldsymbol{1}_n\boldsymbol{1}_n'|.$$

Applying the result in Exercise 14 with $\boldsymbol{A} = -(b/a)\boldsymbol{1}_n$, $\boldsymbol{B} = \boldsymbol{1}_n'$, we get

$$\begin{aligned} a^n |\boldsymbol{I}_n + (b/a)\boldsymbol{1}_n\boldsymbol{1}_n'| &= a^n(1 - \boldsymbol{1}_n'(-b/a\boldsymbol{1}_n)) \\ &= a^{n-1}(a + bn). \end{aligned}$$

CHAPTER 4

4.1 These results can be obtained by direct evaluation.

4.2 Obtained by direct evaluation.

4.3 Obtained by direct evaluation.

4.4 Obtained by direct evaluation.

4.5 **(a)** $\boldsymbol{B}$ must be of order $c \times r$.

(b) If $\boldsymbol{A}$ is $m \times n$, then $\boldsymbol{V}$ must be $n \times n$. Also, if $\boldsymbol{B}$ is $p \times q$, then $\boldsymbol{BVB}'$ implies that $q = n$. Since $\boldsymbol{BVB}'$ is equal to $\boldsymbol{AVA}'$, then $p = m$, that is, $\boldsymbol{B}$ is $m \times n$.

4.6 $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ must be square of the same order. They are not expected to be equal. $tr(\boldsymbol{ABC}) = 6$, $tr(\boldsymbol{BAC}) = 14$.

4.7 This results from direct multiplication. No, the expressions are not, in general, equal to $(\boldsymbol{A}+\boldsymbol{B})'(\boldsymbol{A}+\boldsymbol{B})$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ must have the same order in order for $\boldsymbol{A}+\boldsymbol{B}$ to exist.

4.8 The trace, being the sum of diagonal elements of a square matrix, implies that the trace of each of $\boldsymbol{AA}'$ and $\boldsymbol{A}'\boldsymbol{A}$ is equal to the sum of squares of the elements of $\boldsymbol{A}$. Thus, if such a trace is equal to zero, then this is equivalent to having all the elements of $\boldsymbol{A}$ equal to zero.

4.9 **(a)** $\boldsymbol{AB}$ and $\boldsymbol{BA}$ must be equal.

(b) $tr(\boldsymbol{AB}) = tr(\boldsymbol{B}'\boldsymbol{A}') = tr(\boldsymbol{B}'\boldsymbol{A}) = tr(\boldsymbol{AB}')$.

4.10 This follows by direct substitution.

4.11 Follows by direct multiplication.

4.12 We have

$$[\boldsymbol{y}' - (\boldsymbol{x}'\boldsymbol{y}/\boldsymbol{x}'\boldsymbol{x})\boldsymbol{x}'][\boldsymbol{y} - (\boldsymbol{x}'\boldsymbol{y}/\boldsymbol{x}'\boldsymbol{x})\boldsymbol{x}] = \boldsymbol{y}'\boldsymbol{y} - (\boldsymbol{x}'\boldsymbol{y})^2/\boldsymbol{x}'\boldsymbol{x} \geq 0.$$

This implies $(\sum_{i=1}^n x_i y_i)^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2$. If r is a product-moment correlation, then

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right]^{1/2} \left[\sum_{i=1}^n (y_i - \bar{y})^2\right]^{1/2}}.$$

This implies that $|r| \leq 1$.

4.13 It is sufficient to show that (b) implies (c). Part (b) implies that

$$2tr(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}) - 2tr(\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}'\boldsymbol{A}) = 0,$$

which implies

$$tr[(\boldsymbol{A}'\boldsymbol{B} - \boldsymbol{B}'\boldsymbol{A})(\boldsymbol{A}'\boldsymbol{B} - \boldsymbol{B}'\boldsymbol{A})'] = 0.$$

This indicates that $\boldsymbol{A}'\boldsymbol{B} = \boldsymbol{B}'\boldsymbol{A}$.

4.14 **(a)** This is trivial.

(b) Let a_{ij} and b_{ij} be the (i,j)th elements of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Then,

$$\sum_{i=1}^{n} \boldsymbol{u}_i'\boldsymbol{A}\boldsymbol{B}\boldsymbol{u}_i = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji},$$

which is equal to the sum of the diagonal elements of $\boldsymbol{AB}$. Furthermore,

$$\sum_{i=1}^{n} \boldsymbol{u}_i'\boldsymbol{B}\boldsymbol{A}\boldsymbol{u}_i = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij},$$

which is the sum of the diagonal elements of $\boldsymbol{BA}$. This is the same as the sum of the diagonal elements of $\boldsymbol{AB}$.

4.15 $(\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{A}-\boldsymbol{B}$ implies $\boldsymbol{A}^2 - \boldsymbol{AB} - \boldsymbol{BA} + \boldsymbol{B}^2 = \boldsymbol{A} - \boldsymbol{B}$, hence, $\boldsymbol{A} - \boldsymbol{AB} - \boldsymbol{BA} + \boldsymbol{B} = \boldsymbol{A} - \boldsymbol{B}$, that is, $2\boldsymbol{B} = \boldsymbol{AB} + \boldsymbol{BA}$. Multiply this equality on the left by $\boldsymbol{A}$, get $\boldsymbol{AB} + \boldsymbol{ABA} = 2\boldsymbol{AB}$, which gives $\boldsymbol{AB} = \boldsymbol{ABA}$. Now, multiply the same equality on the right by $\boldsymbol{A}$, get $\boldsymbol{ABA} + \boldsymbol{BA} = 2\boldsymbol{BA}$, which gives $\boldsymbol{ABA} = \boldsymbol{BA}$. We then have $\boldsymbol{AB} = \boldsymbol{BA}$. From $2\boldsymbol{B} = \boldsymbol{AB} + \boldsymbol{BA}$ we conclude $2\boldsymbol{B} = 2\boldsymbol{AB} = 2\boldsymbol{BA}$.

4.16 **(a)** $(\boldsymbol{A} + \lambda\boldsymbol{1}\boldsymbol{1}')\boldsymbol{x} = \boldsymbol{0}$. Since $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{A} + \lambda\boldsymbol{1}\boldsymbol{1}'$ must be singular.

(b) $(\boldsymbol{A} + \boldsymbol{f}\boldsymbol{1}')\boldsymbol{x} = \boldsymbol{0}$. Since $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{A} + \boldsymbol{f}\boldsymbol{1}'$ must be singular.

4.17 $\boldsymbol{X}'\boldsymbol{X}$ is $p \times p$ of rank p, hence, it is nonsingular. It follows that $\boldsymbol{X}\boldsymbol{X}'\boldsymbol{X}$ is of rank p.

4.18 **(a)** The equation can be written as $(\boldsymbol{X} + \boldsymbol{I}_n)^2 = \boldsymbol{0}$. Hence, $|\boldsymbol{X} + \boldsymbol{I}_n|^2 = 0$ which implies that $\boldsymbol{X} + \boldsymbol{I}_n$ is singular.

(b) and **(c)** Writing the same equation as $\boldsymbol{X}^2 + 2\boldsymbol{X} = -\boldsymbol{I}_n$ we conclude that $|\boldsymbol{X}||\boldsymbol{X} + 2\boldsymbol{I}_n| = (-1)^n$. This indicates that both $\boldsymbol{X}$ and $\boldsymbol{X} + 2\boldsymbol{I}_n$ are nonsingular.

(c) Multiplying the equation on the left by $\boldsymbol{X}^{-1}$, we get $\boldsymbol{X} + 2\boldsymbol{I}_n + \boldsymbol{X}^{-1} = \boldsymbol{0}$. Hence, $\boldsymbol{X}^{-1} = -(\boldsymbol{X} + 2\boldsymbol{I}_n)$.

4.19 Let $\boldsymbol{e} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$, where $\boldsymbol{v}'\boldsymbol{v} = \| \boldsymbol{v} \|^2$. Then, $\boldsymbol{v}\boldsymbol{v}' - \boldsymbol{v}'\boldsymbol{v}\boldsymbol{I} = \boldsymbol{v}'\boldsymbol{v}(\boldsymbol{e}\boldsymbol{e}' - \boldsymbol{I})$. Let $\boldsymbol{A} = \boldsymbol{e}\boldsymbol{e}' - \boldsymbol{I}$. Define $\boldsymbol{E}$ to be a matrix of order $n \times (n-1)$ whose columns are orthonormal and orthogonal to $\boldsymbol{e}$. Then,

$$\boldsymbol{A} = \boldsymbol{e}\boldsymbol{e}' - \boldsymbol{I} = [\boldsymbol{e} : \boldsymbol{E}]\text{diag}(1, \boldsymbol{O}_{n-1})[\boldsymbol{e} : \boldsymbol{E}]' - \boldsymbol{I}.$$

Since $[\boldsymbol{e} : \boldsymbol{E}]'[\boldsymbol{e} : \boldsymbol{E}] = \boldsymbol{I}$, $[\boldsymbol{e} : \boldsymbol{E}]' = [\boldsymbol{e} : \boldsymbol{E}]^{-1}$, and we get $[\boldsymbol{e} : \boldsymbol{E}][\boldsymbol{e} : \boldsymbol{E}]' = \boldsymbol{I}$. We then have

$$|\boldsymbol{A}| = |[\boldsymbol{e} : \boldsymbol{E}][\mathrm{diag}(1, \boldsymbol{O}_{n-1}) - \boldsymbol{I}][\boldsymbol{e} : \boldsymbol{E}]'|.$$

The right-hand side is equal to $|\boldsymbol{e} : \boldsymbol{E}|^2 |diag(0, -1, -1, \dots, -1)| = 0$. We conclude that $|\boldsymbol{A}| = 0$.

4.20 Straightforward computations.

4.21 Straightforward computations.

4.22 Straightforward computations.

4.23 To show $\boldsymbol{H}^2 = \boldsymbol{I}$. This is true since $(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}')(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}') = \boldsymbol{I}$.

CHAPTER 5

5.1 This sum is easy to check that it is symmetric.

5.2 (b) and (d) are undefined. (c) is quadratic, (e) is bilinear, (g) is bilinear, and (h) is quadratic.

5.3 If $\boldsymbol{A} = -\boldsymbol{A}'$, then $a_{ji} = -a_{ij}$ for $i \neq j$. We also have $a_{ii} = -a_{ii}$, that is, $a_{ii} = 0$ for all i.

Consider now $\boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{I})\boldsymbol{x}$, which is equal to

$$\begin{aligned}
\boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{I})\boldsymbol{x} &= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{x} \\
&= \frac{1}{2}[\boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}] + \boldsymbol{x}'\boldsymbol{x} \\
&= \boldsymbol{x}'\boldsymbol{x} > 0, \quad \text{for all } \boldsymbol{x} \neq \boldsymbol{0}.
\end{aligned}$$

5.4 $\boldsymbol{A} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}') + \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}')$.

5.5 $(\boldsymbol{A}\boldsymbol{B})' = \boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{B}$.

5.6 $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}$ implies that $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}' = \boldsymbol{X}$. It also follows that $\boldsymbol{X}^2 = \boldsymbol{X}$.

5.7 **(a)** $\boldsymbol{X}^2 = \boldsymbol{0}$ implies that $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{0}$ and hence $tr(\boldsymbol{X}'\boldsymbol{X}) = 0$. Thus, $\sum_{i=1}^{n} \boldsymbol{x}_i'\boldsymbol{x}_i = 0$. The ith column of $\boldsymbol{X}$ is zero for all i. We conclude that $\boldsymbol{X} = \boldsymbol{0}$.

(b) Consider the following matrix, which we denote by $\boldsymbol{Q}$:

$$\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}' - \boldsymbol{X}')(\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}).$$

We then have

$$\boldsymbol{Q} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}'\boldsymbol{X}$$

We note that the right-hand side of the equality is equal to zero. We can then write

$$tr[(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}' - \boldsymbol{X}')(\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X})] = 0.$$

We conclude that $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}$.

5.8 We have

$$\begin{aligned} tr[\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'] &= tr[\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}] \\ &= tr(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}). \end{aligned}$$

Let $\boldsymbol{Y}$ be defined as $\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ and let its kth column be denoted by $\boldsymbol{y}_k$, that is, $\boldsymbol{y}_k = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{x}_k$, $k = 1, 2, \ldots, c$. Then,

$$\begin{aligned} tr[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})] &= \sum_{k=1}^{c} \boldsymbol{y}_k'\boldsymbol{y}_k \\ &= \sum_{k=1}^{c} (\boldsymbol{x}_k'\boldsymbol{A}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{x}_k) \\ &= \sum_{k=1}^{c}\sum_{j=1}^{c} (\boldsymbol{x}_k'\boldsymbol{A}\boldsymbol{x}_j)(\boldsymbol{x}_j'\boldsymbol{A}\boldsymbol{x}_k) \\ &= \sum_{j=1}^{c}\sum_{k=1}^{c} (\boldsymbol{x}_k'\boldsymbol{A}\boldsymbol{x}_j)^2. \end{aligned}$$

5.9 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two orthogonal matrices. Then, $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$ and $\boldsymbol{B}\boldsymbol{B}' = \boldsymbol{I}$. It follows that $(\boldsymbol{A}\boldsymbol{B})(\boldsymbol{B}'\boldsymbol{A}') = \boldsymbol{I}$, that is, $\boldsymbol{A}\boldsymbol{B}$ is orthogonal.

5.10 **(a)** We have $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{A} = \boldsymbol{A}'$. Hence, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} \geq 0$.

(b) We have $\boldsymbol{X}^2 = \boldsymbol{X}$ and $\boldsymbol{Y}^2 = \boldsymbol{Y}$. Hence, $(\boldsymbol{X}\boldsymbol{Y})(\boldsymbol{X}\boldsymbol{Y}) = \boldsymbol{X}^2\boldsymbol{Y}^2 = \boldsymbol{X}\boldsymbol{Y}$.

(c) $\boldsymbol{x}'(\boldsymbol{I} + \boldsymbol{K}\boldsymbol{K}')\boldsymbol{x} \geq \boldsymbol{x}'\boldsymbol{x} > 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}$.

5.11 **(a)** Choose $\boldsymbol{A} = \boldsymbol{I}$.

(b) Choose $\boldsymbol{B} = \boldsymbol{1}\boldsymbol{1}'/5$.

(c) Let $\boldsymbol{y} = \boldsymbol{x} - 5\boldsymbol{1}$. The given quantity is equal to $\boldsymbol{y}'\boldsymbol{y}$. Since $5 = \frac{1}{5}\boldsymbol{1}'\boldsymbol{x}$, then $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{1}(\boldsymbol{1}'\boldsymbol{x}/5) = \left(\boldsymbol{I} - \frac{\boldsymbol{J}}{5}\right)\boldsymbol{x}$, where $\boldsymbol{J} = \boldsymbol{1}\boldsymbol{1}'$. Hence, $\boldsymbol{y}'\boldsymbol{y} = \boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{J}/5)\boldsymbol{x}$, which implies that $\boldsymbol{C} = \boldsymbol{I} - \frac{1}{5}\boldsymbol{J}$.

5.12 **(a)** True since $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is equal to its transpose.

(b) True since $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ is a scalar.

(c) $\boldsymbol{x}'\boldsymbol{C}\boldsymbol{x} = tr(\boldsymbol{x}'\boldsymbol{C}\boldsymbol{x}) = tr(\boldsymbol{C}\boldsymbol{x}\boldsymbol{x}')$.

5.13 **(a)** Let $\boldsymbol{P}$ be the matrix

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then,

$$\boldsymbol{P}'\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

But,

$$\boldsymbol{P}\boldsymbol{P}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(b) We have $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}_n$, hence, $\boldsymbol{P}$ is nonsingular since the determinant of $\boldsymbol{P}$ is equal to one. Now, from $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}_n$ we get $\boldsymbol{P}\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{P}$ by a multiplication of the two sides on the right by $\boldsymbol{P}$. Multiplying now both sides on the left by $\boldsymbol{P}^{-1}$ we get $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_n$.

5.14 **(a)** If $\boldsymbol{AB} = \boldsymbol{B}$, then $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{AB} - \boldsymbol{BA} + \boldsymbol{B}$. But, $\boldsymbol{AB} = \boldsymbol{B} = \boldsymbol{BA}$. This implies that $\boldsymbol{A} - \boldsymbol{B}$ is idempotent. Hence, $\boldsymbol{x}'(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{A} - \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x} \geq 0$, that is, $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

(b) We have

$$tr[(\boldsymbol{AB} - \boldsymbol{B})'(\boldsymbol{AB} - \boldsymbol{B})] = tr[B(-\boldsymbol{A} + \boldsymbol{B})\boldsymbol{B}].$$

The right-hand side is nonpositive since $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite, hence, $\boldsymbol{B}(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{B}$ is positive semidefinite and its eigenvalues are nonnegative, hence $tr[\boldsymbol{B}(-\boldsymbol{A} + \boldsymbol{B})\boldsymbol{B}] \leq 0$.

On the other hand, $tr[(\boldsymbol{AB} - \boldsymbol{B})'(\boldsymbol{AB} - \boldsymbol{B})] \geq 0$ since if $\boldsymbol{D} = \boldsymbol{AB} - \boldsymbol{B}$, then $tr(\boldsymbol{D}'\boldsymbol{D}) = \sum_{i=1}^{n} \boldsymbol{d}_i'\boldsymbol{d}_i \geq 0$, where $\boldsymbol{d}_i$ is the ith column of $\boldsymbol{D}$. We therefore conclude that $tr[(\boldsymbol{AB} - \boldsymbol{B})'(\boldsymbol{AB} - \boldsymbol{B})] = 0$, which implies that $\boldsymbol{AB} - \boldsymbol{B} = \boldsymbol{0}$.

(c) If $\boldsymbol{AB} = \boldsymbol{B}$, then

$$\begin{aligned}(\boldsymbol{A} - \boldsymbol{B})^2 &= \boldsymbol{A} - \boldsymbol{AB} - \boldsymbol{BA} + \boldsymbol{B} \\ &= \boldsymbol{A} - 2\boldsymbol{B} + \boldsymbol{B} \\ &= \boldsymbol{A} - \boldsymbol{B}.\end{aligned}$$

Vice versa, if $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{B}$, then $\boldsymbol{A} - \boldsymbol{AB} - \boldsymbol{BA} + \boldsymbol{B} = \boldsymbol{A} - \boldsymbol{B}$. This gives, $-\boldsymbol{AB} - \boldsymbol{BA} + 2\boldsymbol{B} = \boldsymbol{0}$, that is,

$$(\boldsymbol{B} - \boldsymbol{BA}) + (\boldsymbol{B} - \boldsymbol{AB}) = \boldsymbol{0}.$$

Multiply the last equality on the left by $\boldsymbol{B}$, we get

$$(\boldsymbol{B} - \boldsymbol{BA}) + (\boldsymbol{B} - \boldsymbol{BAB}) = \boldsymbol{0}.$$

Thus, $\boldsymbol{BAB} = 2\boldsymbol{B} - \boldsymbol{BA} = \boldsymbol{AB}$. Hence, $\boldsymbol{BA} = \boldsymbol{BAB}$. We also have

$$(\boldsymbol{B} - \boldsymbol{BAB}) + (\boldsymbol{B} - \boldsymbol{BAB}) = \boldsymbol{0},$$

Thus, $\boldsymbol{B} = \boldsymbol{BAB}$ which finally implies $\boldsymbol{B} = \boldsymbol{AB}$.

5.15 Let $\boldsymbol{a}_j$ and $\boldsymbol{b}_j$ be the jth columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, $j = 1, 2, \ldots, n$. Then,

$$\begin{aligned}
tr(\boldsymbol{A}'\boldsymbol{A}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2 \\
tr(\boldsymbol{B}'\boldsymbol{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij}^2 \\
tr(\boldsymbol{A}'\boldsymbol{B}) &= \sum_{i=1}^{n} \boldsymbol{a}_i'\boldsymbol{b}_i \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij},
\end{aligned}$$

By the Cauchy–Schwarz inequality,

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij}\right)^2 \leq \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij}^2\right).$$

5.16 Let $\boldsymbol{C} = \boldsymbol{AB}$, then $\boldsymbol{C}' = -\boldsymbol{BA}$. Since $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$, we get $tr(\boldsymbol{C}) = tr(-\boldsymbol{C}') = -tr(\boldsymbol{C}') = -tr(\boldsymbol{C})$, which implies $tr(\boldsymbol{C}) = 0$.

5.17 Let $\boldsymbol{B} = \boldsymbol{A} - \boldsymbol{A}'$. Then,

$$\begin{aligned}
tr(\boldsymbol{B}'\boldsymbol{B}) &= tr[(\boldsymbol{A}' - \boldsymbol{A})(\boldsymbol{A} - \boldsymbol{A}')] \\
&= tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}] - tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}'] \\
&= tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}] - tr[\boldsymbol{A}(\boldsymbol{A} - \boldsymbol{A}')] \\
&= tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}] + tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}] \\
&= 2tr[(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{A}] = 0,
\end{aligned}$$

since $\boldsymbol{A}' - \boldsymbol{A}$ is skew-symmetric. It follows that $\boldsymbol{B} = \boldsymbol{0}$ and hence, $\boldsymbol{A} = \boldsymbol{A}'$.

5.18 $\boldsymbol{x}'\boldsymbol{P}'\boldsymbol{APx} = \boldsymbol{y}'\boldsymbol{Ay}$, where $\boldsymbol{y} = \boldsymbol{Px}$. We have $\boldsymbol{y}'\boldsymbol{Ay} > 0$ and is equal to zero only when $\boldsymbol{y} = \boldsymbol{0}$, that is, $\boldsymbol{x} = \boldsymbol{0}$. Hence, $\boldsymbol{P}'\boldsymbol{AP}$ is positive definite.
If $\boldsymbol{A}$ is positive semidefinite, then $\boldsymbol{y}'\boldsymbol{Ay} \geq 0$ and can be zero for some $\boldsymbol{y} \neq \boldsymbol{0}$, that is, there exists a nonzero $\boldsymbol{x}$ such that $\boldsymbol{x}'\boldsymbol{P}'\boldsymbol{APx} = 0$. Hence, $\boldsymbol{P}'\boldsymbol{AP}$ is positive semidefinite.

5.19 **(a)** We have $|\boldsymbol{V}| = |\boldsymbol{P}'||\boldsymbol{P}| = |\boldsymbol{P}|^2$. This implies that $\boldsymbol{P}$ is nonsingular since $|\boldsymbol{V}| \neq 0$.

(b) **(b)** If $\boldsymbol{AV}$ is idempotent, then $(\boldsymbol{AP'P})(\boldsymbol{AP'P}) = \boldsymbol{AP'P}$. Multiplying both sides on the right by the inverse of $\boldsymbol{P'P}$, we get $\boldsymbol{AP'PA} = \boldsymbol{A}$. Then, $(\boldsymbol{PAP'})(\boldsymbol{PAP'}) = \boldsymbol{PAP'}$. This shows that $\boldsymbol{PAP'}$ is idempotent.
Vice versa, if $\boldsymbol{PAP'}$ is idempotent, then $(\boldsymbol{PAP'})(\boldsymbol{PAP'}) = \boldsymbol{PAP'}$. Multiplying both sides on the left by the inverse of $\boldsymbol{P}$ and on the right by the inverse of $\boldsymbol{P}'$, we get $\boldsymbol{AP'PA} = \boldsymbol{A}$. This gives $\boldsymbol{AVA} = \boldsymbol{A}$. Multiplying both sides on the right by $\boldsymbol{V}$, we get $\boldsymbol{AVAV} = \boldsymbol{AV}$ which indicates that $\boldsymbol{AV}$ is idempotent.

(c) If $\boldsymbol{AV}$ is idempotent, then $\boldsymbol{PAP'}$ is idempotent by part (b). Hence, $rank(\boldsymbol{PAP'}) = tr(\boldsymbol{PAP'}) = tr(\boldsymbol{AP'P}) = tr(\boldsymbol{AV})$.

5.20 Suppose that (i) and (ii) are true, then $\boldsymbol{P}' = \boldsymbol{P}^{-1}$ which implies $\boldsymbol{PP}' = \boldsymbol{PP}^{-1} = \boldsymbol{I}_n$. Suppose that (i) and (iii) are true, then $\boldsymbol{P}' = \boldsymbol{P}^{-1}$ which implies $\boldsymbol{P'P} = \boldsymbol{P}^{-1}\boldsymbol{P} = \boldsymbol{I}_n$. If (ii) and (iii) are true, then $\boldsymbol{P'P} = \boldsymbol{I}_n$ and $\boldsymbol{PP}' = \boldsymbol{I}_n$. If $\boldsymbol{P}$ is of order $m \times n$, then $\boldsymbol{P'P} = \boldsymbol{I}_n$ and $\boldsymbol{PP}' = \boldsymbol{I}_n$. This implies that $m = n$.

5.21 The matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, and $\boldsymbol{P}_3$ are given by

$$\boldsymbol{P}_1 = \frac{1}{6}\boldsymbol{I}_4 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_2 - \frac{1}{24}\boldsymbol{J}_4 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_2$$
$$\boldsymbol{P}_2 = \frac{1}{8}\boldsymbol{J}_4 \otimes \boldsymbol{I}_3 \otimes \boldsymbol{J}_2 - \frac{1}{24}\boldsymbol{J}_4 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_2$$
$$\boldsymbol{P}_3 = \frac{1}{2}\boldsymbol{I}_4 \otimes \boldsymbol{I}_3 \otimes \boldsymbol{J}_2 - \frac{1}{6}\boldsymbol{I}_4 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_2 - \frac{1}{8}\boldsymbol{J}_4 \otimes \boldsymbol{I}_3 \otimes \boldsymbol{J}_2 + \frac{1}{24}\boldsymbol{J}_4 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_2.$$

5.22 We have that

$$\left(\boldsymbol{A} - \frac{1}{c}\boldsymbol{dd}'\right)(\boldsymbol{A}^{-1} + \gamma \boldsymbol{A}^{-1}\boldsymbol{dd}'\boldsymbol{A}^{-1})$$
$$= \boldsymbol{I} + \gamma \boldsymbol{dd}'\boldsymbol{A}^{-1} - \frac{1}{c}\,\boldsymbol{dd}'\boldsymbol{A}^{-1} - \frac{\gamma}{c}(\boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})\boldsymbol{dd}'\boldsymbol{A}^{-1}$$
$$= \boldsymbol{I} + (\boldsymbol{dd}'\boldsymbol{A}^{-1})/(c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d}) - \frac{1}{c}\,\boldsymbol{dd}'\boldsymbol{A}^{-1} - [(\boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})/(c(c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})]\boldsymbol{dd}'\boldsymbol{A}^{-1}$$
$$= \boldsymbol{I} + [c\boldsymbol{dd}'\boldsymbol{A}^{-1} - (c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})\boldsymbol{dd}'\boldsymbol{A}^{-1} - (\boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})\boldsymbol{dd}'\boldsymbol{A}^{-1}]/[c(c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})]$$
$$= \boldsymbol{I}.$$

We note that γ exists since $c > \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d}$ and $c > 0$ since $\boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d} = \boldsymbol{d}'\boldsymbol{A}^{-1/2}\boldsymbol{A}^{-1/2}\boldsymbol{d} > 0$.

5.23 **(a)** $\boldsymbol{X}'\boldsymbol{X}$ is nonsingular if $\boldsymbol{X}$ is of full column rank, and $\boldsymbol{x}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{x} \geq 0$ if $\boldsymbol{X}\boldsymbol{x} = \boldsymbol{0}$ which implies $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{x} = \boldsymbol{0}$. Hence, $\boldsymbol{x} = \boldsymbol{0}$.

(b) $\boldsymbol{YY}'$ is nonsingular if $\boldsymbol{Y}$ is of full row rank, and $\boldsymbol{x}'\boldsymbol{YY}'\boldsymbol{x} \geq 0$ if $\boldsymbol{Y}'\boldsymbol{x} = \boldsymbol{0}$. Hence, $\boldsymbol{YY}'\boldsymbol{x} = \boldsymbol{0}$ which implies $\boldsymbol{x} = \boldsymbol{0}$.

CHAPTER 6

6.1 The computations here are straightforward.

6.2 We have $\boldsymbol{Ax} = \lambda\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$. Hence, $(\boldsymbol{A}^2 + \boldsymbol{A})\boldsymbol{x} = (\lambda^2 + \lambda)\boldsymbol{x}$. This implies that $\lambda^2 + \lambda$ is an eigenvalue of $\boldsymbol{A}^2 + \boldsymbol{A}$.

6.3 Let x_1 and x_2 be elements of an eigenvector of the given matrix. We then have $(a - \lambda)x_2 + bx_2 = 0$. One solution is given by $x_1 = -b$, $x_2 = a - \lambda$. The characteristic equation when $a = d$ is $\lambda^2 - 2a\lambda + a^2 - bc = 0$. Hence, the sum of the eigenvalues is equal to $-(-2a)/1 = 2a$.

6.4 The characteristic equation gives

$$(a - \lambda)[(a - \lambda - b)(a - \lambda + b)] - b^2(a - \lambda - b) + b^2(b - a + \lambda) = 0$$

We have one root, $\lambda = a - b$. The other two roots are $\lambda = a - b$ and $\lambda = a + 2b$.

6.5 **(a)** $\boldsymbol{x}'\boldsymbol{x}$ is the only nonzero eigenvalue of $\boldsymbol{xx}'$. By the spectral decomposition theorem,

$$\boldsymbol{xx}' = \boldsymbol{P}\text{diag}(\boldsymbol{x}'\boldsymbol{x}, \boldsymbol{D})\boldsymbol{P}',$$

where $\boldsymbol{D}$ is a diagonal matrix of order $(n-1) \times (n-1)$ of zero elements and $\boldsymbol{P}$ is an orthogonal matrix of eigenvectors. The characteristic equation is $|\boldsymbol{I} - \boldsymbol{xx}' - \lambda\boldsymbol{I}| = 0$, or alternatively,

$$|\boldsymbol{P}[(1 - \lambda)\boldsymbol{I} - \text{diag}(\boldsymbol{x}'\boldsymbol{x}, \boldsymbol{D})]\boldsymbol{P}'| = 0.$$

The roots of this equation are $1 - \boldsymbol{x}'\boldsymbol{x}$ of multiplicity one and 1 of multiplicity $n - 1$.

(b) $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{I} - \boldsymbol{xx}'$ for the eigenvalue $1 - \boldsymbol{x}'\boldsymbol{x}$. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n-1}$ form a basis for the $(n-1)$-dim Euclidean space that are orthogonal to $\boldsymbol{x}$, where the $\boldsymbol{u}_i$'s are unit vectors. These correspond to eigenvectors of $\boldsymbol{I} - \boldsymbol{xx}'$ corresponding to the eigenvalue 1 since $(\boldsymbol{I} - \boldsymbol{xx}')\boldsymbol{u}_i = \boldsymbol{u}_i$ for $i = 1, 2, \ldots, n-1$. Hence, $\boldsymbol{P} = [\boldsymbol{x}* : \boldsymbol{u}_1 : \boldsymbol{u}_2 : \ldots : \boldsymbol{u}_{n-1}]$, where $\boldsymbol{x}* = \boldsymbol{x}/ \parallel \boldsymbol{x} \parallel$, and $\parallel \boldsymbol{x} \parallel$ is the Euclidean norm of $\boldsymbol{x}$.

(c) We have

$$\boldsymbol{PP}' = \frac{\boldsymbol{xx}'}{(\parallel \boldsymbol{x} \parallel)^2} + \sum_{i=1}^{n-1} \boldsymbol{u}_i\boldsymbol{u}_i'.$$

Since $\boldsymbol{PP}' = \boldsymbol{I}$, we get $\sum_{i=1}^{n-1} \boldsymbol{u}_i\boldsymbol{u}_i' = \boldsymbol{I} - \boldsymbol{xx}'/(\parallel \boldsymbol{x} \parallel)^2$. Hence,

$$\begin{aligned}
\boldsymbol{P}\text{diag}(1 - \boldsymbol{x}'\boldsymbol{x}, \boldsymbol{I}_{n-1})\boldsymbol{P}' &= (1 - \boldsymbol{x}'\boldsymbol{x})\frac{\boldsymbol{xx}'}{(\parallel \boldsymbol{x} \parallel)^2} + \sum_{i=1}^{n-1} \boldsymbol{u}_i\boldsymbol{u}_i' \\
&= (1 - \boldsymbol{x}'\boldsymbol{x})\frac{\boldsymbol{xx}'}{(\parallel \boldsymbol{x} \parallel)^2} + \boldsymbol{I} - \frac{\boldsymbol{xx}'}{(\parallel \boldsymbol{x} \parallel)^2} \\
&= \boldsymbol{I} - \boldsymbol{xx}'.
\end{aligned}$$

6.6 We have that $1 + \lambda$ is an eigenvalue of $\boldsymbol{I} + \boldsymbol{A}$. Hence,

$$|\boldsymbol{I} + \boldsymbol{A} - (1 + \lambda)\boldsymbol{I}| = 0$$
$$|\boldsymbol{I} - (1 + \lambda)(\boldsymbol{I} + \boldsymbol{A})^{-1}| = 0$$
$$|(\boldsymbol{I} + \boldsymbol{A})^{-1} - \frac{1}{1 + \lambda}\boldsymbol{I}| = 0.$$

Thus, $1/(1 + \lambda)$ is an eigenvalue of $(\boldsymbol{I} + \boldsymbol{A})^{-1}$.

6.7 Let λ be an eigenvalue of $\boldsymbol{A}$. Then, $(\lambda - 1)^2 \geq 0$, that is, $\lambda^2 - 2\lambda + 1 \geq 0$. Since $\lambda > 0$, then $\lambda + \frac{1}{\lambda} \geq 2$. But, $\lambda + \frac{1}{\lambda}$ is an eigenvalue of $\boldsymbol{A} + \boldsymbol{A}^{-1}$.

6.8 $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 = tr(\boldsymbol{A}'\boldsymbol{A})$, where a_{ij} is the (j, i)th element of $\boldsymbol{A}$. We also have $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix of corresponding eigenvectors. Hence, since $\boldsymbol{A}' = \boldsymbol{A}$, $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}^2 = \boldsymbol{P}\boldsymbol{\Lambda}^2\boldsymbol{P}'$. It follows that $tr(\boldsymbol{A}'\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i^2$.

6.9 By Theorem 6.4, if λ is an eigenvalue of an orthogonal matrix, then $\lambda = \pm 1$. Hence, $1/\lambda$ is an eigenvalue of $\boldsymbol{A}$.

6.10 By Theorem 6.12, the eigenvalues of $\boldsymbol{ABA}$ are the same as the eigenvalues of $\boldsymbol{BAA}$, which is equal to $\boldsymbol{BA}$ since $\boldsymbol{A}$ is idempotent.

6.11 We have that $\boldsymbol{Tx} = \lambda\boldsymbol{x}$. Then, $\boldsymbol{TKx} = \boldsymbol{KTx} = \lambda\boldsymbol{Kx}$. Hence, $\boldsymbol{Kx}$ is an eigenvector of $\boldsymbol{T}$.

6.12 **(a)** We have that $|\boldsymbol{A} - \lambda\boldsymbol{V}| = 0$. Hence, $|\boldsymbol{V}^{-1/2}||\boldsymbol{A} - \lambda\boldsymbol{V}||\boldsymbol{V}^{-1/2}| = 0$. It follows that $|\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2} - \lambda\boldsymbol{I}| = 0$. Hence, λ is an eigenvalue of $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$.

(b) Let $\boldsymbol{x}$ be an eigenvector of $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$ with an eigenvalue λ. Then,

$$\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}\boldsymbol{x} = \lambda\boldsymbol{x}.$$

Let $\boldsymbol{t} = \boldsymbol{V}^{-1/2}\boldsymbol{x}$, then, $\boldsymbol{x} = \boldsymbol{V}^{1/2}\boldsymbol{t}$. We have $\boldsymbol{V}^{-1/2}\boldsymbol{At} = \lambda\boldsymbol{V}^{1/2}\boldsymbol{t}$. Hence, $\boldsymbol{At} = \lambda\boldsymbol{Vt}$ which implies that $\boldsymbol{t}$ is an eigenvector of $\boldsymbol{A}$ with respect to $\boldsymbol{V}$. We thus have $\boldsymbol{t} = \boldsymbol{V}^{-1/2}\boldsymbol{x}$, where $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$ with respect to λ.

6.13 **(a)** To show that $\boldsymbol{A}^k = \boldsymbol{A}$ for all k. The statement is true for $k = 1$. Suppose that it is for $k - 1$, to show it is true for k: $\boldsymbol{A}^k = \boldsymbol{A}\boldsymbol{A}^{k-1} = \boldsymbol{AA} = \boldsymbol{A}$.

(b) This is true since $\boldsymbol{A}^k = \boldsymbol{A}$.

(c) We have $(\boldsymbol{I} - \boldsymbol{A})^2 = \boldsymbol{I} - 2\boldsymbol{A} + \boldsymbol{A}^2 = \boldsymbol{I} - \boldsymbol{A}$. Hence, $\boldsymbol{I} - \boldsymbol{A}$ is idempotent. It follows that $r(\boldsymbol{I} - \boldsymbol{A}) = tr(\boldsymbol{I} - \boldsymbol{A}) = n - tr(\boldsymbol{A}) = n - r(\boldsymbol{A})$.

6.14 We have that $\boldsymbol{Q} = \boldsymbol{RR}'$. The eigenvalues of $\boldsymbol{YQ}$ are the same as those of $\boldsymbol{QY} = \boldsymbol{RR}'\boldsymbol{Y}$, and the eigenvalues of $\boldsymbol{RR}'\boldsymbol{Y}$ are the same as those of $\boldsymbol{R}'\boldsymbol{YR}$.

6.15 The matrix $\boldsymbol{J}$ has one nonzero eigenvalue equal to r (it is the same as the nonzero eigenvalue of $\boldsymbol{1}'\boldsymbol{1} = r$) and a corresponding eigenvector equal to $\boldsymbol{1}$ (since $\boldsymbol{J1} = r\boldsymbol{1}$). Hence, there is an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{J} = \boldsymbol{P}\mathrm{diag}(r, 0, 0, \ldots, 0)\boldsymbol{P}'$. It follows that $a\boldsymbol{I} + b\boldsymbol{J} = \boldsymbol{P}\mathrm{diag}(a + br, a, a, \ldots, a)\boldsymbol{P}'$. Thus the eigenvalues of $a\boldsymbol{I} + b\boldsymbol{J}$ are $a + br$ and a of multiplicity $r - 1$. The corresponding eigenvectors (the columns of

$\boldsymbol{P}$) are $(1/\sqrt{r})\mathbf{1}$, and the remaining ones form an orthonormal basis of the columns $\boldsymbol{I}-(1/r)\boldsymbol{J}$ since $\boldsymbol{I}-(1/r)\boldsymbol{J}$ is of rank $r-1$ and $\mathbf{1}'[\boldsymbol{I}-(1/r)\boldsymbol{J}]=\mathbf{1}'-\mathbf{1}'=\mathbf{0}$.

6.16 The eigenvalues of $\boldsymbol{C}^{-1}\boldsymbol{AC}$ are the same as those of $\boldsymbol{ACC}^{-1}=\boldsymbol{A}$. Furthermore, the eigenvalues of $\boldsymbol{CAC}^{-1}$ are the same as those of $\boldsymbol{AC}^{-1}\boldsymbol{C}=\boldsymbol{A}$.

6.17 **(a)** By the spectral decomposition theorem, we have $\boldsymbol{A}=\boldsymbol{P\Lambda P}'$ where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$, and $\boldsymbol{P}$ is an orthogonal matrix whose columns are eigenvectors of $\boldsymbol{A}$. If $\boldsymbol{\Lambda}=\mathbf{0}$, then $\boldsymbol{A}=\mathbf{0}$.

(b) Let $\boldsymbol{A}$ be given by

$$\boldsymbol{A}=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}.$$

The eigenvalues of $\boldsymbol{A}$ are zero since $|\boldsymbol{A}-\lambda\boldsymbol{I}|=\lambda^2=0$. Hence, $\lambda=0$, but $\boldsymbol{A}\neq\mathbf{0}$.

6.18 **(a)** By the spectral decomposition theorem, we have $\boldsymbol{A}=\boldsymbol{P\Lambda P}'$. We also have that $\boldsymbol{\Lambda}\leq\lambda_{max}\boldsymbol{I}$ and $\boldsymbol{\Lambda}\geq\lambda_{min}\boldsymbol{I}$, where $\leq$ means that for the matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, $\boldsymbol{A}_1\leq\boldsymbol{A}_2$ if and only if $\boldsymbol{A}_2-\boldsymbol{A}_1$ is non-negative. Hence, $\boldsymbol{A}\leq\lambda_{max}\boldsymbol{I}$ and $\boldsymbol{A}\geq\lambda_{min}\boldsymbol{I}$. Thus, for any $\boldsymbol{x}\neq\mathbf{0}$,

$$\frac{\lambda_{min}\boldsymbol{x}'\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\leq\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\leq\frac{\lambda_{max}\boldsymbol{x}'\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}},$$

which gives the desired result.

(b) If $\boldsymbol{x}_{max}$ is an eigenvector of $\boldsymbol{A}$ for the eigenvalue λ_{max}, then $\boldsymbol{x}'_{max}\boldsymbol{A}\boldsymbol{x}_{max}=\lambda_{max}\boldsymbol{x}'_{max}\boldsymbol{x}_{max}$. Hence,

$$\frac{\boldsymbol{x}'_{max}\boldsymbol{A}\boldsymbol{x}_{max}}{\boldsymbol{x}'_{max}\boldsymbol{x}_{max}}=\lambda_{max}.$$

Similarly, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}=\lambda_{min}$ if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ corresponding to the eigenvalue λ_{min}.

6.19 We have that $\sum_{ij}a_{ij}=\mathbf{1}'_n\boldsymbol{A}\mathbf{1}_n$. Hence,

$$\frac{1}{n}\lambda_{min}\mathbf{1}'_n\mathbf{1}_n\leq\frac{1}{n}\mathbf{1}'_n\boldsymbol{A}\mathbf{1}_n\leq\frac{1}{n}\lambda_{max}\mathbf{1}'_n\mathbf{1}_n.$$

It follows that

$$\lambda_{min}\leq\frac{1}{n}\mathbf{1}'_n\boldsymbol{A}\mathbf{1}_n\leq\lambda_{max}.$$

6.20 By the spectral decomposition theorem, we have $\boldsymbol{A}=\boldsymbol{P\Lambda P}'$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix of eigenvectors. Let $\boldsymbol{p}_i$ denote the ith column of $\boldsymbol{P}$, $i=1,2,\ldots,n$. Then, $\boldsymbol{A}=\sum_{i=1}^n\lambda_i\boldsymbol{p}_i\boldsymbol{p}'_i$. Let $\boldsymbol{A}_i=\boldsymbol{p}_i\boldsymbol{p}'_i$. Then $\boldsymbol{A}_i$ is idempotent of rank 1. Furthermore, for $i\neq j$, $\boldsymbol{A}_i\boldsymbol{A}_j=\boldsymbol{p}_i\boldsymbol{p}'_i\boldsymbol{p}_j\boldsymbol{p}'_j=\mathbf{0}$ since $\boldsymbol{p}'_i\boldsymbol{p}_j=0$.

6.21 Since $\boldsymbol{A}$ is symmetric and non-negative definite, we can write,

$$e_{min}(\boldsymbol{B})\boldsymbol{A} \leq \boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2} \leq e_{max}(\boldsymbol{B})\boldsymbol{A}.$$

Hence,

$$e_{min}(\boldsymbol{B})tr(\boldsymbol{A}) \leq tr(\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2}) \leq e_{max}(\boldsymbol{B})tr(\boldsymbol{A}).$$

This is true since, in general, if $\boldsymbol{C} \leq \boldsymbol{D}$, then $\boldsymbol{D} - \boldsymbol{C}$ is positive semidefinite, hence its eigenvalues are non-negative which implies that $tr(\boldsymbol{D} - \boldsymbol{C}) \geq 0$, that is, $tr(\boldsymbol{C}) \leq tr(\boldsymbol{D})$. From the above double inequality, we conclude that

$$e_{min}(\boldsymbol{B})tr(\boldsymbol{A}) \leq tr(\boldsymbol{B}\boldsymbol{A}) = tr(\boldsymbol{A}\boldsymbol{B}) \leq e_{max}(\boldsymbol{B})tr(\boldsymbol{A}).$$

6.22 Let $\boldsymbol{u}_i$ be an $n \times 1$ vector whose elements are all equal to zero except for the ith element which is equal to one. Then, $a_{ii} = \boldsymbol{u}_i'\boldsymbol{A}\boldsymbol{u}_i$, $i = 1, 2, \ldots, n$. We also have

$$e_{min}(\boldsymbol{A})\boldsymbol{u}_i'\boldsymbol{u}_i \leq \boldsymbol{u}_i'\boldsymbol{A}\boldsymbol{u}_i \leq e_{max}(\boldsymbol{A})\boldsymbol{u}_i'\boldsymbol{u}_i.$$

Since $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$, we conclude

$$e_{min}(\boldsymbol{A}) \leq a_{ii} \leq e_{max}(\boldsymbol{A}).$$

6.23 $|\boldsymbol{A} - \theta\boldsymbol{B}| = 0$ implies $|\boldsymbol{A}\boldsymbol{B}^{-1} - \theta\boldsymbol{I}| = 0$. Hence, θ is an eigenvalue of $\boldsymbol{A}\boldsymbol{B}^{-1}$, which is equal to $\boldsymbol{A}\boldsymbol{B}^{-1/2}\boldsymbol{B}^{-1/2}$. The latter matrix has the same eigenvalues as those of $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}$, which is symmetric. Hence, θ is real.

6.24 We have that $tr(\boldsymbol{A}^2) = \sum_{i=1}^{n} \lambda_i^2$, where λ_i is the ith eigenvalue of $\boldsymbol{A}$. But, $\lambda_i^2 \leq \lambda_i e_{max}(\boldsymbol{A})$ since $\boldsymbol{A}$ is positive semidefinite. Hence,

$$tr(\boldsymbol{A}^2) \leq e_{max}(\boldsymbol{A}) \sum_{i=1}^{n} \lambda_i = e_{max}(\boldsymbol{A})tr(\boldsymbol{A}).$$

6.25 We have that

$$\begin{aligned}(\boldsymbol{A} \uplus \boldsymbol{B})(x \otimes y) &= \boldsymbol{A}\boldsymbol{x} \otimes \boldsymbol{y} + \boldsymbol{x} \otimes \boldsymbol{B}\boldsymbol{y} \\ &= \lambda\boldsymbol{x} \otimes \boldsymbol{y} + \mu\boldsymbol{x} \otimes \boldsymbol{y} \\ &= (\lambda + \mu)\boldsymbol{x} \otimes \boldsymbol{y}.\end{aligned}$$

6.26 The eigenvalues of $\boldsymbol{A}$ are 2, -1 with corresponding eigenvectors, (0.7071068, 0.7071068)$'$, $(0.3713907, 0.9284767)'$, respectively. Furthermore, the eigenvalues of $\boldsymbol{B}$ are 4, 3 with corresponding eigenvectors, $(0.4472136, -0.8944272)'$, $(-0.7071068, 0.7071068)'$. Apply now the results given in Exercise 6.25.

6.27 The full-rank factorization of $\boldsymbol{A}$ looks like the one in formula (4.41) where $\boldsymbol{B} = \boldsymbol{K}$ and $\boldsymbol{C} = \boldsymbol{L}$. The result now follows directly from applying formula (6.48) in Theorem 6.12.

6.28 This result follows from the fact that

$$|A' - \lambda I| = |(A - \lambda I)'| = |A - \lambda I|.$$

6.29 Let $\boldsymbol{u}$ be an eigenvector of $\boldsymbol{A}$ with a corresponding eigenvalue λ, where, $\boldsymbol{u} = \boldsymbol{a} + i\boldsymbol{b}$ and $\lambda = \alpha + i\beta$. Then, $\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u}$. Define $\bar{\boldsymbol{u}} = \boldsymbol{a} - i\boldsymbol{b}$, $\bar{\lambda} = \alpha - i\beta$. Then, $\boldsymbol{A}\bar{\boldsymbol{u}} = \bar{\lambda}\bar{\boldsymbol{u}}$. It follows that

$$\begin{aligned}
(\boldsymbol{A}\bar{\boldsymbol{u}})'\boldsymbol{u} &= (\bar{\lambda}\bar{\boldsymbol{u}})'\boldsymbol{u} \\
-\bar{\boldsymbol{u}}'\boldsymbol{A}\boldsymbol{u} &= \bar{\lambda}\bar{\boldsymbol{u}}'\boldsymbol{u} \\
-\bar{\boldsymbol{u}}'(\lambda\boldsymbol{u}) &= \bar{\lambda}\bar{\boldsymbol{u}}'\boldsymbol{u} \\
-\lambda\bar{\boldsymbol{u}}'\boldsymbol{u} &= \bar{\lambda}\bar{\boldsymbol{u}}'\boldsymbol{u} \\
-\lambda(\|\ \boldsymbol{u}\ \|)^2 &= \bar{\lambda}(\|\ \boldsymbol{u}\ \|)^2.
\end{aligned}$$

We conclude that $-\lambda = \bar{\lambda}$. This implies that either $\lambda = 0$ or $\alpha = 0$, which makes λ an imaginary number.

CHAPTER 7

7.1 We have that $tr(\boldsymbol{A})$ is the sum of all nonzero eigenvalues of $\boldsymbol{A}$. But, the number of these eigenvalues, which are all equal to one, is equal to rank($\boldsymbol{A}$). Hence, rank($\boldsymbol{A}$) = $tr(\boldsymbol{A})$.

7.2 Using the following R statements (see Chapter 17):
A = matrix(**c**(10,2,6,-5,-11,-8),3,2)
svd(**A**)
we get the singular-value decomposition of $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \boldsymbol{u} * \text{diag}(\boldsymbol{d}) * \boldsymbol{v}',$$

where

$$\boldsymbol{u} = \begin{bmatrix} -0.57735 & 0.707107 \\ -0.57735 & -0.707107 \\ -0.57735 & 0.0 \end{bmatrix},$$

$\boldsymbol{d} = (17.32051, 7.07107)$,

$$\boldsymbol{v} = \begin{bmatrix} -0.6 & 0.8 \\ 0.8 & 0.6 \end{bmatrix}.$$

7.3 The rank of $\boldsymbol{A}$ is the same as the rank of $\boldsymbol{A}'\boldsymbol{A}$, which is equal to

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 91 & -139 & 1237 \\ -139 & 476 & -2 \\ 1237 & -2 & 31696 \end{bmatrix}$$

The rank of $\boldsymbol{A}'\boldsymbol{A}$ is the same as the number of its nonzero eigenvalues which is three. These eigenvalues are 31744.3432, 516.6524, 2.0043. Hence, the rank of $\boldsymbol{A}$ is equal to 3.

7.4 **(a)** $\boldsymbol{K}$ is of order 3×3 since $\boldsymbol{K}(1, 2, -3)' = (1, 2, -3)'$.

(b) From $\boldsymbol{K}\mathbf{1} = \mathbf{0}$ we conclude that $\boldsymbol{K}$ is of rank 2 since there is a linear relationship among its columns.

(c) We have that $\boldsymbol{K} = \boldsymbol{K}^3$. Since $\boldsymbol{K}$ is symmetric, we obtain $\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{\Lambda}^3\boldsymbol{P}'$ (see Theorem 6.2). Therefore, the eigenvalues of $\boldsymbol{K}$ are $\lambda = 0$, and $\lambda = \pm 1$. We cannot have $\lambda = 0$, $\lambda = 1$, $\lambda = 1$ since this makes $\lambda = 1$ a double root which indicates that the derivative of $\lambda^3 - \lambda$, namely, $3\lambda^2 - 1$ will be equal to zero at $\lambda = 1$, which is not the case. Hence, the eigenvalues of $\boldsymbol{K}$ are $\lambda = 0, 1, -1$. Consequently, the trace of $\boldsymbol{K}$, being the sum of eigenvalues, is equal to 0.

(d) From part (c), the eigenvalues of $\boldsymbol{K}$ are 0, 1, and -1. Hence, its determinant, being the product of its eigenvalues as seen in Section 6.2.4, is equal to 0.

7.5 **(a)** We have that $\boldsymbol{A}(\boldsymbol{A} + 2\boldsymbol{I}) = -\boldsymbol{I}$. Hence, $|\boldsymbol{A}||\boldsymbol{A} + 2\boldsymbol{I}| = -1$. We conclude that $|\boldsymbol{A}| \neq 0$.

(b) Multiplying both sides of the given equality by $\boldsymbol{A}^{-1}$, we get $\boldsymbol{A} + 2\boldsymbol{I} + \boldsymbol{A}^{-1} = \mathbf{0}$. Hence, $\boldsymbol{A}^{-1} = -(\boldsymbol{A} + 2\boldsymbol{I})$.

7.6 If $\boldsymbol{A}$ is diagonalizable, then there would exist a nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix. Hence, we get $\boldsymbol{P}^{-1}\boldsymbol{A}^n\boldsymbol{P} = \boldsymbol{D}^n$. It follows that $\boldsymbol{P}\boldsymbol{D}^n\boldsymbol{P}^{-1} = \mathbf{0}$, which implies that $\boldsymbol{D}^n = \mathbf{0}$, and hence, $\boldsymbol{D} = \mathbf{0}$. Consequently, $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \mathbf{0}$ and $\boldsymbol{A}$ is therefore equal to zero [see Section 11.4 in Banerjee and Roy (2014)].

7.7 Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then, $\boldsymbol{A}^2 = \mathbf{0}$. By the result of Exercise 6, we conclude that $\boldsymbol{A}$ is not diagonalizable.

7.8 We have that $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{\Lambda}$. We then get $\boldsymbol{A}\boldsymbol{p}_j = \lambda_j\boldsymbol{p}_j$, where $\boldsymbol{p}_j$ is the jth column of $\boldsymbol{P}$, $j = 1, 2, \ldots, n$. This shows that the $\boldsymbol{p}_j$'s are eigenvectors of $\boldsymbol{A}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. Since $\boldsymbol{P}$ is nonsingular, its columns form a set of linearly independent eigenvectors. It follows that if $\boldsymbol{A}$ has linearly independent eigenvectors, then we can use these eigenvectors as the columns of the matrix $\boldsymbol{P}$. The matrix $\boldsymbol{P}$ satisfies the equality $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues of $\boldsymbol{A}$. We then conclude that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$ and $\boldsymbol{A}$ is therefore diagonalizable [see Section 11.4 in Banerjee and Roy (2014)].

7.9 The columns of $\boldsymbol{P}$ are the eigenvectors of the matrix $\boldsymbol{A}$. Thus $\boldsymbol{P}$ is equal to

$$\boldsymbol{P} = \begin{bmatrix} -0.7165709 & -0.6107079 & 0.05662903 \\ 0.2356812 & 0.7540444 & -0.66505613 \\ -0.6564911 & -0.2417703 & 0.74464320 \end{bmatrix}.$$

It can be verified that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 6.845609 & 0 & 0 \\ 0 & 3.087654 & 0 \\ 0 & 0 & -2.933263 \end{bmatrix}$$

The diagonal elements of $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ are the eigenvalues of $\boldsymbol{A}$.

7.10 Suppose that there exits a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{D}\boldsymbol{P}$. We can then write

$$\begin{aligned} \boldsymbol{A} - \lambda \boldsymbol{I}_2 &= \boldsymbol{P}^{-1}\boldsymbol{D}\boldsymbol{P} - \lambda \boldsymbol{I}_2 \\ &= \boldsymbol{P}^{-1}(\boldsymbol{D} - \lambda \boldsymbol{I}_2)\boldsymbol{P}. \end{aligned}$$

It follows that $\boldsymbol{A}$ and $\boldsymbol{D}$ have the same eigenvalues. Since $\boldsymbol{D}$ is diagonal, the diagonal elements are its eigenvalues. Furthermore, it can be verified that $\boldsymbol{A}$ has one eigenvalue equal to 5 of multiplicity 2. We then conclude that $\boldsymbol{D} = 5\boldsymbol{I}_2$. This makes $\boldsymbol{A} = 5\boldsymbol{I}_2$ contrary to the previously given value of $\boldsymbol{A}$. This contradiction leads us to conclude that $\boldsymbol{A}$ is not diagonable.

CHAPTER 8

8.1 To show that (8.2) satisfies (8.1):

(i) We have that $\boldsymbol{A} = \boldsymbol{K}\boldsymbol{L}$. Then, $\boldsymbol{A}\boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{A}\boldsymbol{L}')^{-1}\boldsymbol{K}'\boldsymbol{A} = \boldsymbol{K}\boldsymbol{L}\boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}$. But, $\boldsymbol{L}\boldsymbol{L}'$ and $\boldsymbol{K}'\boldsymbol{K}$ are nonsingular. We then get, $\boldsymbol{K}(\boldsymbol{L}\boldsymbol{L}')(\boldsymbol{L}\boldsymbol{L}')^{-1}(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L} = \boldsymbol{K}\boldsymbol{L} = \boldsymbol{A}$.

(ii) $\boldsymbol{M}\boldsymbol{A}\boldsymbol{M} = \boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{A}\boldsymbol{L}')^{-1}\boldsymbol{K}'(\boldsymbol{K}\boldsymbol{L})\boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{K}' = \boldsymbol{L}'(\boldsymbol{L}\boldsymbol{L}')^{-1}(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}' = \boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{K}' = \boldsymbol{M}$.

(iii) $\boldsymbol{A}\boldsymbol{M} = \boldsymbol{K}\boldsymbol{L}\boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{K}' = \boldsymbol{K}(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}'$, which is symmetric.

(iv) $\boldsymbol{M}\boldsymbol{A} = \boldsymbol{L}'(\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{L} = \boldsymbol{L}'(\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{L}$, which is symmetric.

(8.4) can similarly be shown to satisfy (8.1).

8.2 Making the proper substitution in $\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}$, we get

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix}.$$

But, $\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} = \boldsymbol{A}_{22}$ by the fact that

$$\begin{bmatrix}\boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix} = \boldsymbol{K}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\end{bmatrix},$$

for some matrix $\boldsymbol{K}$, since the rows of $\begin{bmatrix}\boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$ are linearly dependent on the rows of $\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\end{bmatrix}$. We then have $\boldsymbol{A}_{21} = \boldsymbol{K}\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{22} = \boldsymbol{K}\boldsymbol{A}_{12}$. Hence, $\boldsymbol{K} = \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$. It follows that $\boldsymbol{A}_{22} = \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$.

8.3 We have that $\boldsymbol{A}_r^- = \boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-$. Hence,

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}_r^-\boldsymbol{A} &= \boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\\ &= \boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A},\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{A}_r^-\boldsymbol{A}\boldsymbol{A}_r^- &= \boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-\\ &= \boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^-\\ &= \boldsymbol{A}_r^-.\end{aligned}$$

8.4 We have that $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}'$. To show that $\boldsymbol{G}\boldsymbol{X}'$ is a generalized inverse of $\boldsymbol{X}$ that satisfies conditions $(i), (ii), (iii)$ of the Penrose conditions.

(i) To show that $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}$, consider the product, $(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}' - \boldsymbol{X}')(\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X})$, which can easily be shown to be equal to zero, hence its trace is zero. This implies that $\boldsymbol{X}\boldsymbol{G}(\boldsymbol{X}'\boldsymbol{X}) - \boldsymbol{X} = \boldsymbol{0}$. Note that $\boldsymbol{G}$, being a generalized inverse of the symmetric matrix $\boldsymbol{X}'\boldsymbol{X}$, can be chosen to be symmetric.

(ii) To show that $\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}' = \boldsymbol{G}\boldsymbol{X}'$, it is easy to show that the product $(\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}' - \boldsymbol{G}\boldsymbol{X}')(\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G} - \boldsymbol{X}\boldsymbol{G})$ is equal to zero. Hence its trace is zero, which implies that $\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}' - \boldsymbol{G}\boldsymbol{X}' = \boldsymbol{0}$.

(iii) To show that $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'$ is symmetrix: this is trivial considering that $\boldsymbol{G}$ is chosen to be symmetric.

8.5 The Moore–Penrose generalized inverses of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted here by $\boldsymbol{A}^-$, and $\boldsymbol{B}^-$, respectively, are

$$\boldsymbol{A}^- = \begin{bmatrix} 0.03065 & 0.04789 & -0.01341\\ 0.03448 & 0.07471 & 0.00575\\ 0.07280 & 0.00958 & -0.13602\\ -0.09579 & -0.00383 & 0.18774\end{bmatrix},$$

$$\boldsymbol{B}^- = \begin{bmatrix} -1.04672 & 0.96338 & -0.26136 & -0.17803\\ -0.56061 & 0.56061 & -0.13636 & -0.13636\\ 1.04798 & -0.88131 & 0.29454 & 0.12879\\ 0.22601 & -0.30934 & 0.10227 & 0.18561\end{bmatrix}.$$

These generalized inverses were obtained using the function *ginv*(.) in SAS's PROC IML (for details, see Chapter 15).

8.6 **(a)** $\boldsymbol{Q}^{-1}\boldsymbol{A}^{-}\boldsymbol{P}^{-1}$ is a generalized inverse, where $\boldsymbol{A}^{-}$ is a generalized inverse of $\boldsymbol{A}$.

(b) $\boldsymbol{A}^{-}\boldsymbol{A}$ is a generalized inverse.

(c) $\frac{1}{k}\boldsymbol{A}^{-}$ is a generalized inverse.

(d) $\boldsymbol{A}\boldsymbol{B}^{-}\boldsymbol{A}'$ is a generalized inverse.

(e) $\frac{1}{n^2}\boldsymbol{J}$ is a generalized inverse.

8.7 If $\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-}$ is idempotent, then

$$(\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-})(\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-}) = \boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-}.$$

It is easy to show that multiplying both sides on the left by $\boldsymbol{A}$ and on the right by $\boldsymbol{B}$ gives

$$(\boldsymbol{A}\boldsymbol{B})(\boldsymbol{B}^{-}\boldsymbol{A}^{-})(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{A}\boldsymbol{B},$$

which shows that $\boldsymbol{B}^{-}\boldsymbol{A}^{-}$ is a generalized inverse of $\boldsymbol{A}\boldsymbol{B}$.
Vice versa, if the above equality is true, then by multiplying both sides by $\boldsymbol{A}^{-}$ on the left and by $\boldsymbol{B}^{-}$ on the right we can show that $\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-}$ is idempotent.

8.8 $\boldsymbol{A}\boldsymbol{A}_r^{-}\boldsymbol{A} = \boldsymbol{A}$ shows that $r(\boldsymbol{A}) \leq r(\boldsymbol{A}_r^{-})$. Also, $\boldsymbol{A}_r^{-}\boldsymbol{A}\boldsymbol{A}_r^{-} = \boldsymbol{A}_r^{-}$ gives $r(\boldsymbol{A}_r^{-}) \leq r(\boldsymbol{A})$. Hence, $r(\boldsymbol{A}_r^{-}) = r(\boldsymbol{A})$.

8.9 To show that $\boldsymbol{Z}(\boldsymbol{K}\boldsymbol{Z}^{-}\boldsymbol{K})\boldsymbol{Z} = \boldsymbol{Z}$, substituting in $\boldsymbol{Z}\boldsymbol{Z}^{-}\boldsymbol{Z} = \boldsymbol{Z}$, we get $(\boldsymbol{K}\boldsymbol{A}\boldsymbol{K})(\boldsymbol{Z}^{-})(\boldsymbol{K}\boldsymbol{A}\boldsymbol{K}) = \boldsymbol{K}\boldsymbol{A}\boldsymbol{K}$, which can be written as $\boldsymbol{K}\boldsymbol{A}(\boldsymbol{K}\boldsymbol{K})\boldsymbol{Z}^{-}(\boldsymbol{K}\boldsymbol{K})\boldsymbol{A}\boldsymbol{K} = \boldsymbol{K}\boldsymbol{A}\boldsymbol{K}$. This gives, $\boldsymbol{Z}(\boldsymbol{K}\boldsymbol{Z}^{-}\boldsymbol{K})\boldsymbol{Z} = \boldsymbol{Z}$, indicating that $\boldsymbol{K}\boldsymbol{Z}^{-}\boldsymbol{K}$ is a generalized inverse of $\boldsymbol{Z}$.

8.10 To show that $\boldsymbol{A}^2\boldsymbol{G}^2\boldsymbol{A}^2 = \boldsymbol{A}^2$: Since $\boldsymbol{A}$ and $\boldsymbol{G}\boldsymbol{A}$ are symmetric, we get $\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{G}$. Then, $\boldsymbol{A}^2\boldsymbol{G} = \boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}$. Multiplying both sides on the right by $\boldsymbol{G}$, we get, $\boldsymbol{A}^2\boldsymbol{G}^2 = \boldsymbol{A}\boldsymbol{G}$. We next multiply both sides on the right by $\boldsymbol{A}^2$, we get $\boldsymbol{A}^2\boldsymbol{G}^2\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^2$. But, since $\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}$, then $\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^2 = \boldsymbol{A}^2$. Hence, $\boldsymbol{A}^2\boldsymbol{G}^2\boldsymbol{A}^2 = \boldsymbol{A}^2$.

8.11 Let $\boldsymbol{R}$ and $\boldsymbol{S}$ be permutation matrices of orders $m \times m$ and $q \times q$, respectively, such that $\boldsymbol{R}\boldsymbol{P}\boldsymbol{S} = [\boldsymbol{P}_1\ \boldsymbol{P}_2]$, which is of order $m \times q$, where $\boldsymbol{P}_1$ is $m \times m$ of rank m and $\boldsymbol{P}_2$ is of order $m \times (q - m)$ ($q \geq m$ since if $q < m$, then $r(\boldsymbol{P}) \leq q < m$, a contradiction). A generalized inverse of $\boldsymbol{R}\boldsymbol{P}\boldsymbol{S}$ is given by

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{P}_1^{-1} \\ \boldsymbol{0} \end{bmatrix}.$$

Hence, a generalized inverse of $\boldsymbol{P}$ is

$$\boldsymbol{P}^{-} = \boldsymbol{S}\begin{bmatrix} \boldsymbol{P}_1^{-1} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{R}.$$

It follows that

$$\begin{aligned} \boldsymbol{P}\boldsymbol{P}^{-} &= \boldsymbol{R}'[\boldsymbol{P}_1\ \boldsymbol{P}_2]\boldsymbol{S}'\boldsymbol{S}\begin{bmatrix}\boldsymbol{P}_1^{-1}\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{R} \\ &= \boldsymbol{R}'\boldsymbol{I}_m\boldsymbol{R} \\ &= \boldsymbol{I}_m. \end{aligned}$$

8.12 We show first that $\boldsymbol{D} = \boldsymbol{P}'^{-}(\boldsymbol{P}'\boldsymbol{D}\boldsymbol{P})\boldsymbol{P}^{-}$. This is true since by Exercise 8.11, $\boldsymbol{P}\boldsymbol{P}^{-} = \boldsymbol{I}_m$, then $\boldsymbol{P}'^{-}\boldsymbol{P}' = \boldsymbol{I}_m$. Let us now show that $\boldsymbol{D}^{-1} = \boldsymbol{P}(\boldsymbol{P}'\boldsymbol{D}\boldsymbol{P})^{-}\boldsymbol{P}'$. This is true since $\boldsymbol{D}^{-1}\boldsymbol{D} = \boldsymbol{P}(\boldsymbol{P}'\boldsymbol{D}\boldsymbol{P})^{-}\boldsymbol{P}'\boldsymbol{D} = \boldsymbol{P}(\boldsymbol{P}^{-}\boldsymbol{D}^{-1}\boldsymbol{P}'^{-})\boldsymbol{P}'\boldsymbol{D} = \boldsymbol{I}_m$. Furthermore, $\boldsymbol{D}\boldsymbol{D}^{-1} = \boldsymbol{D}\boldsymbol{P}(\boldsymbol{P}'\boldsymbol{D}\boldsymbol{P})^{-}\boldsymbol{P}' = \boldsymbol{D}\boldsymbol{P}(\boldsymbol{P}^{-}\boldsymbol{D}^{-1}\boldsymbol{P}'^{-})\boldsymbol{P}' = \boldsymbol{I}_m$.

8.13 $s = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}$. We now show that $\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y}$. We have that $\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{y}$. Since $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}$ by Theorem 8.2(ii), we get $\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{b}$. Thus, $s = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y}$.

8.14
- **(a)** We have $\boldsymbol{M}\boldsymbol{V}[\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}\boldsymbol{M}]\boldsymbol{M}\boldsymbol{V} = \boldsymbol{M}\boldsymbol{V}[\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}\boldsymbol{M}]\boldsymbol{V}\boldsymbol{M}' = \boldsymbol{M}\boldsymbol{V}\boldsymbol{M}' = \boldsymbol{M}\boldsymbol{M}\boldsymbol{V} = \boldsymbol{M}\boldsymbol{V}$. Hence, $\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}\boldsymbol{M}$ is a generalized inverse of $\boldsymbol{M}\boldsymbol{V}$.
- **(b)** We have $\boldsymbol{M}\boldsymbol{V}(\boldsymbol{V}^{-1}\boldsymbol{M})\boldsymbol{M}\boldsymbol{V} = \boldsymbol{M}^3\boldsymbol{V} = \boldsymbol{M}\boldsymbol{V}$. Hence, $\boldsymbol{V}^{-1}\boldsymbol{M}$ is a generalized inverse of $\boldsymbol{M}\boldsymbol{V}$.
- **(c)** Here, $\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}\boldsymbol{M}\boldsymbol{V} = \boldsymbol{M}\boldsymbol{V}[\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}\boldsymbol{M}]\boldsymbol{M}\boldsymbol{V} = \boldsymbol{M}\boldsymbol{V}$, by part (a). Hence, $\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}')^{-}$ is a generalized inverse of $\boldsymbol{M}\boldsymbol{V}$.

8.15
- **(a)** We have $(\boldsymbol{A}+\boldsymbol{X})\boldsymbol{X} = \boldsymbol{X}$. Hence, $\boldsymbol{X} = (\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{X}$. Multiply both sides on the left by $\boldsymbol{X}$, we get $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{X}$. Hence, $(\boldsymbol{A}+\boldsymbol{X})^{-1}$ is a generalized inverse of $\boldsymbol{X}$.
- **(b)** $\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}(\boldsymbol{A}+\boldsymbol{X})\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X} = \boldsymbol{0}$.
- **(c)** We have $(\boldsymbol{A}+\boldsymbol{X})(\boldsymbol{A}+\boldsymbol{X})^{-1}(\boldsymbol{A}+\boldsymbol{X}) = \boldsymbol{A}+\boldsymbol{X}$. Hence, $\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{A} + +\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{X} + \boldsymbol{X}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{A} + \boldsymbol{X}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{X} = \boldsymbol{A}+\boldsymbol{X}$. Using parts (a) and (b), we conclude that $\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{A} + \boldsymbol{X} = \boldsymbol{A}+\boldsymbol{X}$. It follows that $\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{X})^{-1}\boldsymbol{A} = \boldsymbol{A}$, that is, $(\boldsymbol{A}+\boldsymbol{X})^{-1}$ is a generalized inverse of $\boldsymbol{A}$.

8.16 This follows from the fact that

$$X(X'X)^{-}X' = [X_1\ X_2]\begin{bmatrix}(X_1'X_1)^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}X_1'\\ X_2'\end{bmatrix}.$$

The right-hand side is equal to $\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$.

8.17
- (i) $\boldsymbol{P}$ is symmetric and invariant to $(\boldsymbol{X}'\boldsymbol{X})^{-}$ by Theorem 8.2 (parts (iii) and (iv)). To show it is idempotent, it can be verified that $[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'][\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$.
- (ii) $[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{0}$. Hence, $\boldsymbol{X}'\boldsymbol{P} = \boldsymbol{0}$.
- (iii) Since $\boldsymbol{P}$ is idempotent, then $r(\boldsymbol{P}) = tr[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = n - tr[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = n - r[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = n - r(\boldsymbol{X})$.
- (iv) The columns of $\boldsymbol{P}$ are orthogonal to the columns of $\boldsymbol{X}$ follows from $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{0}$. Hence, $\boldsymbol{0} = \boldsymbol{P}\boldsymbol{x}_i$ for all i, where $\boldsymbol{x}_i$ is the ith column of $\boldsymbol{X}$. Thus, if $\boldsymbol{p}_j$ is the jth column of $\boldsymbol{P}$, then $\boldsymbol{x}_i'\boldsymbol{p}_j = 0, j = 1, 2, \ldots, n$.

(v) Since $r(\boldsymbol{X}) = n - r(\boldsymbol{P})$, then $n = r(\boldsymbol{X}) + r(\boldsymbol{P})$. Thus the columns of $\boldsymbol{X}$ and $\boldsymbol{P}$ span the n-dimensional Euclidean space. Furthermore, the column space of $\boldsymbol{P}$ is the orthogonal complement of the column space of $\boldsymbol{X}$.

(vi) $\boldsymbol{BX} = \boldsymbol{0}$ implies that any row of $\boldsymbol{B}$ is orthogonal to the columns of $\boldsymbol{X}$. Hence, the rows of $\boldsymbol{B}$ are spanned by the rows of $\boldsymbol{P}$. Thus, there is a matrix $\boldsymbol{K}$ such that $\boldsymbol{B} = \boldsymbol{KP}$. It follows that if $\boldsymbol{b}_i'$ is the ith row of $\boldsymbol{B}$, then $\boldsymbol{b}_i' = \boldsymbol{k}_i'\boldsymbol{P}$, where $\boldsymbol{k}_i'$ is the ith row of $\boldsymbol{K}$, that is, $\boldsymbol{b}_i'$ is a linear combination of the rows of $\boldsymbol{P}$.

(vii) We have that $\boldsymbol{X1} = (\boldsymbol{x}'\boldsymbol{1})\boldsymbol{1}_n$, where $\boldsymbol{x}'\boldsymbol{1}$ represents a row sum of $\boldsymbol{X}$. We also have that $\boldsymbol{PX} = \boldsymbol{0}$. Multiplying on the right by $\boldsymbol{1}$, we get $\boldsymbol{PX1} = \boldsymbol{0}$. Thus, $\boldsymbol{P1} = \boldsymbol{0}$, that is, the row sums of $\boldsymbol{P}$ are all zero.

(viii) Since $\boldsymbol{PX} = \boldsymbol{0}$, then $\boldsymbol{P1} = \boldsymbol{0}$, that is, $\boldsymbol{P}$ has zero row sums.

(ix) When $\boldsymbol{X} = \boldsymbol{1}$, then $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \frac{1}{n}\boldsymbol{J}_n$.

(x) If $\boldsymbol{AX} = \boldsymbol{0}$, then $\boldsymbol{A} = \boldsymbol{LP}$ for some matrix $\boldsymbol{L}$ by part (vi). Hence, $\boldsymbol{A} = \boldsymbol{PL}'$ since $\boldsymbol{A}$ is symmetric. Multiplying both sides on the right by $\boldsymbol{P}$, we get $\boldsymbol{AP} = \boldsymbol{PL}'\boldsymbol{P}$. But, $\boldsymbol{A} = \boldsymbol{LP}$. Multiplying both sides on the right by $\boldsymbol{P}$, we get $\boldsymbol{AP} = \boldsymbol{LP}^2 = \boldsymbol{LP} = \boldsymbol{A}$. We then conclude that $\boldsymbol{A} = \boldsymbol{AP} = \boldsymbol{PL}'\boldsymbol{P}$.

8.18 **(a)** $\boldsymbol{X} = \boldsymbol{MT}$ implies $r(\boldsymbol{X}) \le r(\boldsymbol{T})$. Multiplying on the left by $\boldsymbol{M}'$, we get $\boldsymbol{M}'\boldsymbol{X} = \boldsymbol{M}'\boldsymbol{MT}$. Since $\boldsymbol{M}'\boldsymbol{M}$ is nonsingular, we get $\boldsymbol{T} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{X}$. Thus, $r(\boldsymbol{T}) \le r(\boldsymbol{X})$. We conclude that $r(\boldsymbol{T}) = r(\boldsymbol{X})$.

(b) Let $\boldsymbol{M}' = \boldsymbol{Q}$. Then, $\boldsymbol{M}'(\boldsymbol{MM}')^-\boldsymbol{M} = \boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^-\boldsymbol{Q}'$, which is idempotent of rank $= r(\boldsymbol{Q}) = r(\boldsymbol{M}) = p$. But, $\boldsymbol{M}'(\boldsymbol{MM}')^-\boldsymbol{M}$ is of order $p \times p$ and is therefore nonsingular. Hence, it must be equal to the identity.

(c) We have $\boldsymbol{X}'(\boldsymbol{MM}')^-\boldsymbol{X} = \boldsymbol{T}'\boldsymbol{M}'(\boldsymbol{MM}')^-\boldsymbol{MT} = \boldsymbol{T}'\boldsymbol{T}$. Thus, $r[\boldsymbol{X}'(\boldsymbol{MM}')^-\boldsymbol{X}] = r(\boldsymbol{T}'\boldsymbol{T}) = r(\boldsymbol{T}) = r(\boldsymbol{X})$.

8.19 **(a)** We have that $\boldsymbol{A} = \boldsymbol{AA}^-\boldsymbol{A}$. Then, $r(\boldsymbol{A}) \le r(\boldsymbol{A}^-)$.

(b) We have that $r(\boldsymbol{A}) \le r(\boldsymbol{AA}^-)$ and $r(\boldsymbol{A}) \le r(\boldsymbol{A}^-\boldsymbol{A})$. We also have $r(\boldsymbol{AA}^-) \le r(\boldsymbol{A})$ and $r(\boldsymbol{A}^-\boldsymbol{A}) \le r(\boldsymbol{A})$. We conclude that $r(\boldsymbol{A}) = r(\boldsymbol{A}^-\boldsymbol{A}) = r(\boldsymbol{AA}^-)$.

(c) $\boldsymbol{A}^-\boldsymbol{AA}^-\boldsymbol{A} = \boldsymbol{A}^-\boldsymbol{A}$.

8.20 $\boldsymbol{A}^-\boldsymbol{y}$ is a solution to $\boldsymbol{y} = \boldsymbol{Ax}$. Hence, $\boldsymbol{A}\tilde{\boldsymbol{x}} = \boldsymbol{AA}^-\boldsymbol{y} + (\boldsymbol{AA}^-\boldsymbol{A} - \boldsymbol{A})\boldsymbol{z} = \boldsymbol{AA}^-\boldsymbol{y} = \boldsymbol{y}$.

8.21 Suppose that $\boldsymbol{BA}^-\boldsymbol{A} = \boldsymbol{B}$. Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two solutions of $\boldsymbol{Ax} = \boldsymbol{y}$, that is, $\boldsymbol{Ax}_1 = \boldsymbol{y}$ and $\boldsymbol{Ax}_2 = \boldsymbol{y}$. Then, $\boldsymbol{Bx}_1 = \boldsymbol{BA}^-\boldsymbol{Ax}_1 = \boldsymbol{BA}^-\boldsymbol{y} = \boldsymbol{BA}^-\boldsymbol{Ax}_2 = \boldsymbol{Bx}_2$.
Suppose now that $\boldsymbol{Bx}$ is unique. Let $\tilde{\boldsymbol{x}}$ be a solution to $\boldsymbol{Ax} = \boldsymbol{y}$, that is, $\boldsymbol{B}\tilde{\boldsymbol{x}} = \boldsymbol{BA}^-\boldsymbol{y}$. Using now the general solution given in Exercise 8.20, we can write

$$\boldsymbol{B}[\boldsymbol{A}^-\boldsymbol{y} + (\boldsymbol{A}^-\boldsymbol{A} - \boldsymbol{I})\boldsymbol{z}] = \boldsymbol{BA}^-\boldsymbol{y}, \quad \textit{for all } \boldsymbol{z}.$$

This implies $(\boldsymbol{BA}^-\boldsymbol{A} - \boldsymbol{B})\boldsymbol{z} = \boldsymbol{0}$, for all $\boldsymbol{z}$. We conclude that $\boldsymbol{BA}^-\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{0}$.

8.22 Suppose $\boldsymbol{A}^-\boldsymbol{ABB}^-$ is idempotent. Then, $\boldsymbol{A}^-\boldsymbol{ABB}^-\boldsymbol{A}^-\boldsymbol{ABB}^- = \boldsymbol{A}^-\boldsymbol{ABB}^-$. Multiplying both sides on the left by $\boldsymbol{A}$ and on the right by $\boldsymbol{B}$, we get

$$\boldsymbol{AA}^-\boldsymbol{ABB}^-\boldsymbol{A}^-\boldsymbol{ABB}^-\boldsymbol{B} = \boldsymbol{AA}^-\boldsymbol{ABB}^-\boldsymbol{B}.$$

This implies that $\boldsymbol{ABB}^-\boldsymbol{A}^-\boldsymbol{AB} = \boldsymbol{AB}$. We conclude that $\boldsymbol{B}^-\boldsymbol{A}^-$ is a generalized inverse of $\boldsymbol{AB}$.

Now suppose that $\boldsymbol{B}^{-}\boldsymbol{A}^{-}$ is a generalized inverse of $\boldsymbol{AB}$. Then, $\boldsymbol{ABB}^{-}\boldsymbol{A}^{-}\boldsymbol{AB} = \boldsymbol{AB}$. Multiplying both sides on the left by $\boldsymbol{A}^{-}$ and on the right by $\boldsymbol{B}^{-}$, we get $\boldsymbol{A}^{-}\boldsymbol{ABB}^{-}\boldsymbol{A}^{-}\boldsymbol{ABB}^{-} = \boldsymbol{A}^{-}\boldsymbol{ABB}^{-}$. Hence, $\boldsymbol{A}^{-}\boldsymbol{ABB}^{-}$ is idempotent.

8.23 In Henderson and Searle (1980, p. 15), it is reported that a generalized inverse of $\boldsymbol{A} + \boldsymbol{UBV}$ is given by

$$\boldsymbol{G}_3 = \boldsymbol{A}^{-} - \boldsymbol{A}^{-}\boldsymbol{UB}(\boldsymbol{B} + \boldsymbol{BVA}^{-}\boldsymbol{UB})^{-}\boldsymbol{BVA}^{-}.$$

In this exercise, $\boldsymbol{U} = \boldsymbol{C}$, $\boldsymbol{B} = \boldsymbol{I}$, $\boldsymbol{V} = \boldsymbol{C}'$, and $\boldsymbol{A}^{-} = \boldsymbol{A}^{-1}$. By making the substitution we get

$$(\boldsymbol{A} + \boldsymbol{CC}')^{-} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{I} + \boldsymbol{C}'\boldsymbol{A}^{-1}\boldsymbol{C})^{-}\boldsymbol{C}'\boldsymbol{A}^{-1}.$$

8.24 **(a)** $\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{y}$.

(b) $\mathrm{E}(\boldsymbol{c}'\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{\beta}$. Thus, $\boldsymbol{c}'\boldsymbol{X\beta} = \boldsymbol{a}'\boldsymbol{\beta}$. This implies $\boldsymbol{a}' = \boldsymbol{c}'\boldsymbol{X}$, that is, $\boldsymbol{a}'$ belongs to the row space of $\boldsymbol{X}$.

(c) $\boldsymbol{a}'\hat{\boldsymbol{\beta}} = \boldsymbol{c}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$. The result follows from the fact that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is invariant to $(\boldsymbol{X}'\boldsymbol{X})^{-}$ (Theorem 8.2, part (iii)).

CHAPTER 9

9.1 For the Jacobian matrix we have

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} 12x_1x_2 + 2x_2 & 6x_1^2 + 2x_1 + 2x_2 \\ 6x_1^2 + 2x_1 + 2x_2 & 2x_1 \end{bmatrix}.$$

For the Hessian matrix we have

$$\begin{aligned} \boldsymbol{H} &= \left\{ \frac{\partial^2 y}{\partial x_i \partial x_j} \right\} \\ &= \begin{bmatrix} \dfrac{\partial^2 y}{\partial x_1^2} & \dfrac{\partial^2 y}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 y}{\partial x_2 \partial x_1} & \dfrac{\partial^2 y}{\partial x_2^2} \end{bmatrix} \\ &= \begin{bmatrix} 12x_1x_2 + 2x_2 & 6x_1^2 + 2x_1 + 2x_2 \\ 6x_1^2 + 2x_1 + 2x_2 & 2x_1 \end{bmatrix}. \end{aligned}$$

9.2

$$\boldsymbol{A} = \begin{bmatrix} 7 & 2 & 3 \\ 2 & -5 & -3 \\ 3 & -3 & 3 \end{bmatrix}.$$

9.3 Let $\boldsymbol{X\beta} = \boldsymbol{\gamma}$. Then, the given ratio can be written as $\boldsymbol{\gamma}'\boldsymbol{tt}'\boldsymbol{\gamma}/\boldsymbol{\gamma}'\boldsymbol{\gamma}$. The nonzero eigenvalues of $\boldsymbol{tt}'$ are the same as those of $\boldsymbol{t}'\boldsymbol{t}$. Hence, $\boldsymbol{tt}'$ has one nonzero eigenvalue, namely, $\boldsymbol{t}'\boldsymbol{t}$. This maximum is achieved when $\boldsymbol{\gamma} = \boldsymbol{t}/\sqrt{\boldsymbol{t}'\boldsymbol{t}}$, which is a unit eigenvector of $\boldsymbol{tt}'$ for the eigenvalue $\boldsymbol{t}'\boldsymbol{t}$. If $\boldsymbol{\beta}_{max}$ denotes the value of $\boldsymbol{\beta}$ that maximizes the ratio, then $\boldsymbol{X\beta_{max}} = \boldsymbol{\gamma}$. We then conclude that $\boldsymbol{\beta_{max}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{t}/\sqrt{\boldsymbol{t}'\boldsymbol{t}}$.

9.4 $|\boldsymbol{A}|^2 = |\boldsymbol{AA}'| \leq \prod_{i=1}^{n} \boldsymbol{d}_i'\boldsymbol{d}_i$ by Theorem 9.8, where $\boldsymbol{d}_i'$ is the ith row of $\boldsymbol{A}$. Hence, $|\boldsymbol{A}|^2 \leq \prod_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right)$.

9.5 By the spectral decomposition theorem we have $\boldsymbol{A} = \boldsymbol{P\Lambda P}'$. Hence, $\boldsymbol{A}^{1/2} = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{P}'$. Let $\boldsymbol{x}'\boldsymbol{A}^{1/2} = \boldsymbol{z}'$, and $\boldsymbol{A}^{1/2}\boldsymbol{y} = \boldsymbol{w}$. Then, $(\boldsymbol{x}'\boldsymbol{Ay})^2 = (\boldsymbol{z}'\boldsymbol{w})^2 = \left(\sum_{i=1}^{n} w_i z_i\right)^2 \leq \sum_{i=1}^{n} z_i^2 \sum_{i=1}^{n} w_i^2 = (\boldsymbol{x}'\boldsymbol{Ax})(\boldsymbol{y}'\boldsymbol{Ay})$.

9.6 $tr(\boldsymbol{AX}) = \sum_{i=1}^{n} \boldsymbol{a}_i'\boldsymbol{x}_i$, where $\boldsymbol{a}_i'$ is the ith row of $\boldsymbol{A}$ and $\boldsymbol{x}_i$ is the ith column of $\boldsymbol{X}$. It is then easy to see that $tr(\boldsymbol{AX}) = (vec\boldsymbol{A}')'vec\boldsymbol{X}$. Since $tr(\boldsymbol{AX}) = tr(\boldsymbol{XA})$, then $tr(\boldsymbol{XA}) = (vec\boldsymbol{X}')'vec\boldsymbol{A}$.

9.7 Let $\boldsymbol{C} = \boldsymbol{AB} - \boldsymbol{BA}$. Then, $\boldsymbol{CC}' = \boldsymbol{AB}^2\boldsymbol{A} - (\boldsymbol{AB})^2 - (\boldsymbol{BA})^2 + \boldsymbol{BA}^2\boldsymbol{B}$. Since $tr(\boldsymbol{CC}') \geq 0$, then we must have $2tr(\boldsymbol{A}^2\boldsymbol{B}^2) - 2tr(\boldsymbol{AB})^2 \geq 0$, that is, $tr(\boldsymbol{AB})^2 \leq tr(\boldsymbol{A}^2\boldsymbol{B}^2)$. Equality holds if and only if $\boldsymbol{AB} = \boldsymbol{BA}$: If this condition is true, then $tr(\boldsymbol{AB})^2 = tr(\boldsymbol{ABBA}) = tr(\boldsymbol{A}^2\boldsymbol{B}^2)$. Vice versa, if $tr(\boldsymbol{AB})^2 = tr(\boldsymbol{A}^2\boldsymbol{B}^2)$, then $tr(\boldsymbol{CC}') = 0$ which implies that $\boldsymbol{C} = \boldsymbol{0}$. Hence, $\boldsymbol{AB} = \boldsymbol{BA}$.

9.8 The Hessian matrix is

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}\left[\frac{\partial(\boldsymbol{x}'\boldsymbol{Ax})}{\partial \boldsymbol{x}'}\right] &= \frac{\partial}{\partial \boldsymbol{x}}\left[\frac{\partial(\boldsymbol{x}'\boldsymbol{Ax})}{\partial \boldsymbol{x}}\right]' \\
&= \frac{\partial}{\partial \boldsymbol{x}}[2\boldsymbol{Ax}]' \\
&= 2\frac{\partial(\boldsymbol{x}'\boldsymbol{A})}{\partial \boldsymbol{x}} \\
&= 2\boldsymbol{A}.
\end{aligned}
$$

9.9 We have

$$
\begin{aligned}
\frac{\partial(\boldsymbol{a}'\boldsymbol{Wa})}{\partial \boldsymbol{W}} &= \frac{\partial}{\partial \boldsymbol{W}}\left(\sum_{ij} a_i a_j w_{ij}\right) \\
&= \left\{\sum_{ij} a_i a_j \frac{\partial(w_{ij})}{\partial w_{kl}}\right\} \\
&= \left\{a_k a_l\right\} \\
&= \boldsymbol{aa}'.
\end{aligned}
$$

9.10 The (i, k)th element of $\frac{\partial \boldsymbol{z}'}{\partial \boldsymbol{x}}$ is

$$\begin{aligned}\frac{\partial z_k}{\partial x_i} &= \sum_{j=1}^{n} \frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_i}\\ &= \frac{\partial z_k}{\partial \boldsymbol{y}'}\frac{\partial \boldsymbol{y}}{\partial x_i}, \quad i = 1, 2, \dots, m; k = 1, 2, \dots, p.\end{aligned}$$

Hence,

$$\begin{aligned}\frac{\partial \boldsymbol{z}'(\boldsymbol{x})}{\partial \boldsymbol{x}} &= \left\{\frac{\partial z_j}{\partial \boldsymbol{y}'}\frac{\partial \boldsymbol{y}}{\partial x_i}\right\}_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, p\\ &= \left\{\frac{\partial \boldsymbol{y}'}{\partial x_i}\frac{\partial z_j}{\partial \boldsymbol{y}}\right\}_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, p\\ &= \left[\frac{\partial \boldsymbol{y}'}{\partial x_i}\right]_{i=1,\dots,m}\left[\frac{\partial z_j}{\partial \boldsymbol{y}}\right]_{j=1,\dots,p}.\end{aligned}$$

Hence,

$$\frac{\partial \boldsymbol{z}'(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}'}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{z}'}{\partial \boldsymbol{y}}.$$

9.11 By the spectral decomposition theorem, $\boldsymbol{A}^2 = \boldsymbol{P}\boldsymbol{\Lambda}^2\boldsymbol{P}'$. Thus, $tr(\boldsymbol{A}^2) = \sum_{i=1}^{n} \lambda_i^2$. But, $\sum_{i=1}^{n} \lambda_i^2 \leq e_{max}(\boldsymbol{A}) \sum_{i=1}^{n} \lambda_i = e_{max}(\boldsymbol{A})tr(\boldsymbol{A})$.
Equality holds if and only if $\lambda_i = \lambda$ for all i: If $\lambda_i = \lambda$ for all i, then $\boldsymbol{A} = \lambda \boldsymbol{I}_n$. Then, $\boldsymbol{A}^2 = \lambda^2 \boldsymbol{I}_n$, and $tr(\boldsymbol{A}^2) = n\lambda^2 = e_{max}(\boldsymbol{A}) \sum_{i=1}^{n} \lambda_i = n\lambda^2$. Vice versa, if $\sum_{i=1}^{n} \lambda_i^2 = e_{max}(\boldsymbol{A}) \sum_{i=1}^{n} \lambda_i$, then $\sum_{i=1}^{n}(e_{max}(\boldsymbol{A})\lambda_i - \lambda_i^2) = 0$, which implies $e_{max}(\boldsymbol{A})\lambda_i - \lambda_i^2 = 0$ for all i since $e_{max}(\boldsymbol{A})\lambda_i \geq \lambda_i^2$ for all i. We then conclude that $\lambda_i = e_{max}(\boldsymbol{A})$ for all i

9.12 This follows directly from the definition of $\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}}$ in Section 9.5.3.

9.13 We have that $tr[(\boldsymbol{A} - \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})] = tr(\boldsymbol{A}^2) + tr(\boldsymbol{B}^2) - 2tr(\boldsymbol{AB})$. Let λ_i be the ith eigenvalue of $\boldsymbol{A} - \boldsymbol{B}$. Since $\boldsymbol{A} - \boldsymbol{B}$ is symmetric, then $tr(\boldsymbol{A} - \boldsymbol{B})^2 = \sum_{i=1}^{m} \lambda_i^2 \geq 0$. Hence, $tr(\boldsymbol{A}^2) + tr(\boldsymbol{B}^2) - 2tr(\boldsymbol{AB}) \geq 0$, that is, $tr(\boldsymbol{AB}) \leq \frac{1}{2}[tr(\boldsymbol{A}^2) + tr(\boldsymbol{B}^2)]$.

9.14 $tr[\boldsymbol{A} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{B} - 2\boldsymbol{A} \otimes \boldsymbol{B}] = [tr(\boldsymbol{A})]^2 + [tr(\boldsymbol{B})]^2 - 2tr(\boldsymbol{A})tr(\boldsymbol{B}) = [tr(\boldsymbol{A}) - tr(\boldsymbol{B})]^2 \geq 0$. We can conclude that

$$tr(\boldsymbol{A} \otimes \boldsymbol{B}) \leq \frac{1}{2}tr[\boldsymbol{A} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{B}]$$

9.15 Since $\boldsymbol{A}$ is non-negative definite we can write

$$e_{min}(\boldsymbol{B})\boldsymbol{A} \leq \boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2} \leq e_{max}(\boldsymbol{B})\boldsymbol{A}.$$

Hence,

$$e_{min}(\boldsymbol{B})tr(\boldsymbol{A}) \leq tr(\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2}) \leq e_{max}(\boldsymbol{B})tr(\boldsymbol{A}).$$

We conclude that

$$e_{min}(\boldsymbol{B})tr(\boldsymbol{A}) \leq tr(\boldsymbol{A}\boldsymbol{B}) \leq e_{max}(\boldsymbol{B})tr(\boldsymbol{A}).$$

9.16 We have

$$\begin{aligned}
f(\boldsymbol{A}) &= \boldsymbol{P}\text{diag}[f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)]\boldsymbol{P}' \\
&= \boldsymbol{P}\text{diag}\left[\sum_{i=0}^{\infty} a_i\lambda_1^i, \sum_{i=0}^{\infty} a_i\lambda_2^i, \ldots, \sum_{i=0}^{\infty} a_i\lambda_n^i\right]\boldsymbol{P}' \\
&= \boldsymbol{P}\left[\sum_{i=0}^{\infty} a_i\boldsymbol{\Lambda}^i\right]\boldsymbol{P}' \\
&= \sum_{i=0}^{\infty} a_i(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}')^i \\
&= \sum_{i=0}^{\infty} \boldsymbol{a}_i\boldsymbol{A}^i.
\end{aligned}$$

9.17 $sin(\boldsymbol{A}) = \sum_{i=0}^{\infty}(-1)^i\frac{1}{(2i+1)!}\boldsymbol{A}^{2i+1}$.

9.18 $\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a} = \sum_{p=1}^{n}\sum_{q=1}^{n} a_p a_q x_{pq}$. Hence,

$$\begin{aligned}
\frac{\partial(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a})}{\partial x_{ij}} &= \sum_{p=1}^{n}\sum_{q=1}^{n} a_p a_q \frac{\partial x_{pq}}{\partial x_{ij}} \\
&= a_i a_j + a_j a_i \\
&= 2a_i a_j, \quad i \neq j \\
&= a_i^2, \quad i = j.
\end{aligned}$$

Hence,

$$\frac{\partial(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a})}{\partial x_{ij}} = 2a_i a_j - \delta_{ij} a_i a_j,$$

where $\delta_{ij} = 1$ if $i = j$, otherwise it is equal to zero. We conclude that

$$\frac{\partial(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a})}{\partial \boldsymbol{X}} = 2\boldsymbol{a}\boldsymbol{a}' - \boldsymbol{D}_{\boldsymbol{a}\boldsymbol{a}'}.$$

9.19 If $\frac{\partial(\boldsymbol{QX})}{\partial \boldsymbol{X}} = \boldsymbol{0}$, then $\frac{\partial(\boldsymbol{QX})}{\partial x_{ij}} = \boldsymbol{0}$ for all $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$. We have that $\frac{\partial(\boldsymbol{QX})}{\partial x_{ij}} = \boldsymbol{Q}E_{ij} = \boldsymbol{Q}\boldsymbol{e}_i\boldsymbol{e}_j' = \boldsymbol{0}$. Multiplying both sides on the right by $\boldsymbol{e}_j$ and noting that $\boldsymbol{e}_j'\boldsymbol{e}_j = 1$, we get $\boldsymbol{Q}\boldsymbol{e}_i = \boldsymbol{0}$ for all $i = 1, 2, \ldots, p$. Hence,

$$\boldsymbol{Q}[\boldsymbol{e}_1\ \boldsymbol{e}_2\ \ldots\ \boldsymbol{e}_p] = \boldsymbol{0}.$$

Since $[\boldsymbol{e}_1\ \boldsymbol{e}_2\ \ldots\ \boldsymbol{e}_p] = \boldsymbol{I}_p$, we conclude that $\boldsymbol{Q} = \boldsymbol{0}$.

9.20 Since $|\lambda_i| < 1$, the power series $\sum_{j=0}^{\infty} \lambda_i^j$ is absolutely convergent and its sum is equal to $\frac{1}{1-\lambda_i}$, $i = 1, 2, \ldots, n$. Hence, by the result of Exercise 9.16, we have $\sum_{j=0}^{\infty} \boldsymbol{A}^j = (\boldsymbol{I}_n - \boldsymbol{A})^{-1}$.

9.21 By Theorem 9.8, we have $|\boldsymbol{A}_j| \leq \prod_{i=1}^{j} a_{ii}$, $j = 1, 2, \ldots, n$, since $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ are positive definite (the leading principal minors of $\boldsymbol{A}$ are all positive). Furthermore, since the leading principal minors are positive, then it can be shown by induction and by applying Theorem 9.8 that $a_{ii} > 0$ for $i = 1, 2, \ldots, n$. Hence, by Carleman's inequality, $\sum_{j=1}^{n}(a_{11}a_{22}\ldots a_{jj})^{1/j} \leq (1+\frac{1}{n})^n \sum_{j=1}^{n} a_{jj}$. It follows that

$$\begin{aligned}\sum_{j=1}^{n} |\boldsymbol{A}_j|^{1/j} &\leq \sum_{j=1}^{n}(a_{11}a_{22}\ldots a_{jj})^{1/j} \\ &\leq \left(1+\frac{1}{n}\right)^n \sum_{j=1}^{n} a_{jj} \\ &= \left(1+\frac{1}{n}\right)^n tr(\boldsymbol{A}).\end{aligned}$$

CHAPTER 10

10.1 $var(\boldsymbol{y}) = \boldsymbol{T}var(\boldsymbol{x})\boldsymbol{T}' = \boldsymbol{T}'\boldsymbol{T} = \boldsymbol{I}$.

10.2 We have $\boldsymbol{S} = \boldsymbol{X}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{X}$. Let $\boldsymbol{X}^* = \boldsymbol{X} - \theta\boldsymbol{J}_n$. Then,

$$\begin{aligned}(\boldsymbol{X}^*)'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)(\boldsymbol{X}^*) &= (\boldsymbol{X} - \theta\boldsymbol{J}_n)'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)(\boldsymbol{X} - \theta\boldsymbol{J}_n) \\ &= \boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X} - \theta\boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{J}_n \\ &\quad -\theta\boldsymbol{J}_n\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X} + \theta^2\boldsymbol{J}_n\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{J}_n \\ &= \boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{X}.\end{aligned}$$

10.3 **(a)** $SS_A = n\sum_{i=1}^{a}(\bar{x}_{i.})^2 - an(\bar{x}_{..})^2$. Hence,

$$SS_A = \boldsymbol{x}\left[\frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n\right]\boldsymbol{x}.$$

(b) $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n} x_{ij}^2 - n\sum_{i=1}^{a}(\bar{x}_{i.})^2$. Hence,

$$SS_E = \boldsymbol{x}'\left[\boldsymbol{I}_a \otimes \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n\right]\boldsymbol{x}.$$

(c) From the expressions for SS_A and SS_E, the product of their corresponding matrices, namely,

$$\left[\frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n\right]\left[\boldsymbol{I}_a \otimes \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n\right]$$

is equal to $\frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n + \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n$, which is equal to the zero matrix. Hence, SS_A and SS_E are independently distributed.

(d) $\frac{1}{\sigma^2}[\frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n](\sigma^2\boldsymbol{I}_{an}) = \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n$ is an idempotent matrix of rank $= tr[\frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n - \frac{1}{an}\boldsymbol{J}_a \otimes \boldsymbol{J}_n] = a-1$. Hence, $SS_A/\sigma^2 \sim \chi^2_{a-1}$ (the noncentrality parameter is equal to zero).
As for SS_E/σ^2, we have that the matrix

$$\frac{1}{\sigma^2}\left[\boldsymbol{I}_a \otimes \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n\right]\left(\sigma^2\boldsymbol{I}_{an}\right) = \boldsymbol{I}_a \otimes \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{I}_a \otimes \boldsymbol{J}_n,$$

is idempotent of rank $= an - a = a(n-1)$ and a noncentrality parameter equal to zero. Hence, $SS_E/\sigma^2 \sim \chi^2_{a(n-1)}$.

10.4 **(a)** $E(\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}) = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}\boldsymbol{\beta} + tr(\boldsymbol{M}\sigma^2\boldsymbol{I}) = \sigma^2 tr(\boldsymbol{M}) = \sigma^2(n - tr\{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\}) = \sigma^2[n - r\{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\}) = \sigma^2(n-r)$.

(b) $\boldsymbol{M}(\sigma^2\boldsymbol{I})(\boldsymbol{I}-\boldsymbol{M}) = \boldsymbol{0}$. Hence, $\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}$ and $\boldsymbol{y}'$ and $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}$ are independent.

(c) $\frac{1}{\sigma^2}\boldsymbol{M}(\sigma^2\boldsymbol{I}) = \boldsymbol{M}$ is idempotent of rank $n-r$. The noncentrality parameter is zero since $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}\boldsymbol{\beta} = 0$. Hence, $\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y} \sim \chi^2_{n-r}$.

(d) $\frac{1}{\sigma^2}(\boldsymbol{I}-\boldsymbol{M})(\sigma^2\boldsymbol{I}) = \boldsymbol{I}-\boldsymbol{M}$ is idempotent of rank $= r[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'] = r$, and the noncentrality parameter is zero since $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0$.

10.5 **(a)** It is legitimate since it is symmetric and positive definite (the eigenvalues, namely, 16.8775, 2.0351, 0.0873 are all positive).

(b) $var(X+2Y+3Z) = var[(1,2,3)var[X\ Y\ Z](1,2,3)'] = 5$, $var(Y+4Z) = var[(0,1,4)var[X\ Y\ Z](0,1,4)'] = 2$. We also have

$$\begin{aligned} cov(X+2Y+3Z, Y+4Z) &= cov[(1,2,3)\boldsymbol{W}, \boldsymbol{W}'(0,1,4)'] \\ &= (1,2,3)var[X\ Y\ Z](0,1,4)' \\ &= 0, \end{aligned}$$

where $\boldsymbol{W} = (X,Y,Z)'$.

10.6 Since $\boldsymbol{V}$ is positive definite, then so is the matrix $\boldsymbol{R}$. Applying Theorem 9.8 to $\boldsymbol{R}$, we get $|\boldsymbol{R}| \leq \prod_{i=1}^{n} r_{ii} \leq 1$ since the diagonal elements of $\boldsymbol{R}$, $r_{11}, r_{22}, \ldots, r_{nn}$, are equal to one.

10.7 We have that $\boldsymbol{S} = \boldsymbol{X}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{X}$. Then, $\boldsymbol{t}'\boldsymbol{X}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{X}\boldsymbol{t} = \boldsymbol{u}'\boldsymbol{u}$, where $\boldsymbol{u} = (\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{X}\boldsymbol{t}$. Since $\boldsymbol{u}'\boldsymbol{u} \geq 0$, we conclude that $\boldsymbol{S}$ is non-negative definite.
An unbiased estimator of $\boldsymbol{V}$ is $\hat{\boldsymbol{V}} = \frac{1}{n-1}\boldsymbol{S}$. An estimated correlation matrix is $\hat{\boldsymbol{R}} = \boldsymbol{D}_s^{-1/2}\hat{\boldsymbol{V}}\boldsymbol{D}_s^{-1/2}$, where $\boldsymbol{D}_s = \mathrm{diag}(s_{11}, s_{22}, \ldots, s_{nn})$, and $s_{jj} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_{.j})^2$. Hence, since $\hat{\boldsymbol{R}}$ is positive definite, then by Theorem 9.8, $|\hat{\boldsymbol{R}}| \leq 1$, given that the diagonal elements of $\hat{\boldsymbol{R}}$ are equal to 1.

10.8 If $\boldsymbol{A} = \boldsymbol{V}^{-1}$, then $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}_n$ is idempotent of rank n. Hence, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ has the chi-squared distribution with n degrees of freedom. Vice versa, if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ has the chi-squared distribution with n degrees of freedom, then $\boldsymbol{A}\boldsymbol{V}$ must be idempotent of rank n. This implies that $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}_n$, that is, $\boldsymbol{A} = \boldsymbol{V}^{-1}$.

10.9 **(a)** We have that

$$\begin{aligned}\frac{\partial Q}{\partial \boldsymbol{x}} &= 2\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \\ &= (4x_1 + 2x_2 - 2x_3 - 6, 2x_1 + 6x_2 - 4x_3 - 6, -2x_1 - 4x_2 + 8x_3 + 10)'.\end{aligned}$$

Furthermore,

$$\frac{\partial[2(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{V}^{-1}]}{\partial \boldsymbol{x}} = 2\boldsymbol{V}^{-1}.$$

Hence,

$$\boldsymbol{V}^{-1} = \frac{1}{2}\begin{bmatrix} 4 & 2 & -2 \\ 2 & 6 & -4 \\ 2 & -4 & 8 \end{bmatrix}.$$

Thus,

$$\boldsymbol{V} = \frac{1}{13}\begin{bmatrix} 8 & -2 & 1 \\ -2 & 7 & 3 \\ 1 & 3 & 5 \end{bmatrix}$$

Also, since $\mathrm{E}[\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})] = \boldsymbol{0}$, we get

$$\begin{bmatrix} 4 & 2 & -2 \\ 2 & 6 & -4 \\ -2 & -4 & 8 \end{bmatrix}\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 6 \\ -10 \end{bmatrix}.$$

Hence, $\boldsymbol{\mu} = (1, 0, -1)'$.

(b) The moment generating function of $\boldsymbol{x}$ is

$$\phi_x(\boldsymbol{t}) = exp\left[\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{V}\boldsymbol{t}\right],$$

where $\boldsymbol{\mu}$ and $\boldsymbol{V}$ are given in part(a).

10.10 $\boldsymbol{x}$ must be normally distributed with mean $= (1, -1, 2)'$ and variance-covariance matrix

$$\boldsymbol{V} = \begin{bmatrix} 2 & -0.5 & -1 \\ -0.5 & 1 & 0 \\ -1 & 0 & 4 \end{bmatrix}.$$

Let $u = (1, 1, -1)\boldsymbol{x}$. Hence, $u \sim N[(1, 1, -1)\boldsymbol{\mu}, (1, 1, -1)\boldsymbol{V}(1, 1, -1)'] = N(-2, 8)$. It follows that

$$\begin{aligned} P(x_1 + x_2 > x_3) &= P(u > 0) \\ &= P\left(z > \frac{2}{\sqrt{8}}\right) \\ &= P(z > 0.707) \\ &= 0.239. \end{aligned}$$

10.11 **(a)** $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y}$ can be expressed as

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y} = \frac{1}{2}\boldsymbol{W}'\boldsymbol{B}\boldsymbol{W},$$

where $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}' & \boldsymbol{0} \end{bmatrix}$ is a symmetric matrix. Then, according to formula (10.33),

$$\mathrm{E}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{\mu_1}'\boldsymbol{A}\boldsymbol{\mu_2} + tr(\boldsymbol{A}\boldsymbol{V}'_{12}),$$

$\boldsymbol{V}'_{12}$ is a portion of the matrix

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}'_{12} & \boldsymbol{V}_{22} \end{bmatrix},$$

and $\boldsymbol{\mu} = (\boldsymbol{\mu_1}', \boldsymbol{\mu_2}')'$.

(b) Using formulas (10.35) and (10.36) we get

$$\begin{aligned} var(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y}) &= \frac{2}{4}tr(\boldsymbol{B}\boldsymbol{V})^2 + \boldsymbol{\mu}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}\boldsymbol{\mu} \\ &= tr(\boldsymbol{A}\boldsymbol{V}'_{12}\boldsymbol{A}\boldsymbol{V}'_{12}) + tr(\boldsymbol{A}\boldsymbol{V}_{22}\boldsymbol{A}'\boldsymbol{V}_{11}) + \boldsymbol{\mu_1}'\boldsymbol{A}\boldsymbol{V}_{22}\boldsymbol{A}'\boldsymbol{\mu_1} + \\ &\quad \boldsymbol{\mu_2}'\boldsymbol{A}'\boldsymbol{V}_{11}\boldsymbol{A}\boldsymbol{\mu_2} + 2\boldsymbol{\mu_1}'\boldsymbol{A}\boldsymbol{V}'_{12}\boldsymbol{A}\boldsymbol{\mu_2}. \end{aligned}$$

10.12 We have

$$var(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}) = var(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) + 2cov(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}) + var(\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}).$$

Hence,

$$cov(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}) = \frac{1}{2}\{var[\boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}] - var(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) - var(\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x})\}.$$

We can then apply Corollary 10.2 to each of the three variances. Doing so, we get

$$cov(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}) = 2tr(\boldsymbol{AVBV}) + 4\boldsymbol{\mu}'\boldsymbol{AVB}\boldsymbol{\mu}.$$

CHAPTER 11

11.1 We have that $\frac{SS_E}{\sigma^2} \sim \chi^2_{n-p}$. Then, the expected value of the reciprocal of MS_E is equal to $(n-p)/\sigma^2$ times the expected value of the reciprocal of χ^2_{n-p}. Let $u = \frac{1}{\chi^2_{n-p}}$, and let $f(x)$ be the density function of χ^2_{n-p}, that is,

$$f(x) = \frac{x^{\frac{n-p}{2}-1}}{\Gamma\left(\frac{n-p}{2}\right)} \frac{e^{-x/2}}{2^{(n-p)/2}}.$$

Then, the expected value of the reciprocal of MS_E is equal to

$$\frac{n-p}{\sigma^2}\int_0^\infty \frac{1}{x} f(x)dx.$$

It is easy to show that

$$\frac{n-p}{\sigma^2}\int_0^\infty \frac{1}{x} f(x)dx = \frac{n-p}{\sigma^2(n-p-2)}.$$

11.2 $\mathrm{E}\left[\frac{n-4}{(n-2)MS_E}\right] = \left(\frac{n-4}{n-2}\right)\left(\frac{n-2}{\sigma^2(n-4)}\right)$ by Exercise 11.1. We then have

$$\mathrm{E}\left[\frac{n-4}{(n-2)MS_E}\right] = \frac{1}{\sigma^2}.$$

11.3 $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$. Its variance is

$$var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right),$$

where $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. Taking the partial derivative of $var(\hat{\beta}_0)$ with respect to x_i, $i = 1, 2, \ldots, n$, and equating it to zero, we get

$$2\bar{x}S_{xx}/n - \bar{x}^2(2x_i - 2n\bar{x}/n = 0.$$

Either $\bar{x} = 0$ or

$$2S_{xx}/n - \bar{x}(2x_i - 2\bar{x}) = 0.$$

This can be written as

$$\frac{2}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}^2 = 2\bar{x}x_i - 2\bar{x}^2.$$

We then have

$$x_i = \frac{\sum_{i=1}^{n} x_i^2}{n\bar{x}}.$$

Averaging over i, we get

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i^2}{n\bar{x}}.$$

This implies that $S_{xx} = 0$, which is not possible. We conclude that $\bar{x} = 0$. The minimum variance of $\hat{\beta}_0$ is equal to $\frac{\sigma^2}{n}$.

11.4 **(a)** The least-squares estimates of the parameters are: $\hat{\beta}_0 = 7.8881$, $\hat{\beta}_1 = 11.0737$, $\hat{\beta}_2 = -4.1303$, and the error mean square is $MS_E = 3.4904$.

(b) The F-statistic for testing the null hypothesis H_0 is given by formula (11.61) where $\boldsymbol{K}$ is the matrix

$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 & -2 \\ 1 & 0 & 0 \end{bmatrix},$$

$\boldsymbol{m}$ is the column vector $\boldsymbol{m} = (0, 8)'$, and $r = 2$. The matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is given by

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 4.1835 & 2.6575 & -3.6068 \\ 2.6575 & 4.1327 & -2.9809 \\ -3.6068 & -2.9809 & 3.3444 \end{bmatrix}.$$

Under H_0, F has the F-distribution with 2 and 9 degrees of freedom. Making the substitution in formula (11.61), we get $F = 41.2866$, which is significant at the $\alpha = 0.05$ level ($F_{0.05,2,9} = 4.26$). Under H_a, $\boldsymbol{m} = (1, 10)'$ and F has the noncentral distribution with 2 and 9 degrees of freedom and a noncentrality parameter given by

$$\theta = \frac{1}{\sigma^2}(\boldsymbol{m}_a - \boldsymbol{m})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{m}_a - \boldsymbol{m}).$$

Hence, $\theta = 22.0249$, and the power value is $P(F > F_{0.05,2,9}|F \sim F_{2,9}(\theta)] = 0.9481$.

(c) The 95% confidence interval on β_i is given by $\hat{\beta}_i \pm t_{0.025,9}(a_{ii}MS_E)^{1/2}$, where a_{ii} is the ith diagonal element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Hence, the 95% confidence intervals for β_0, β_1, and β_3 are $(-0.7556, 16.5318)$, $(2.4826, 19.6648)$, $(-11, 8587, 3.5981)$, respectively.

11.5 We have that $\beta_0 + \beta_1 x_0 = 0$. Let $\hat{\Delta} = \hat{\beta}_0 + \hat{\beta}_1 x_0$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least-squares estimates of β_0 and β_1, respectively. Then, $\hat{\Delta}$ is normally distributed with a zero mean and a variance given by $var(\hat{\Delta}) = \boldsymbol{z}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}_0\sigma^2$, where $\boldsymbol{z}_0 = (1, x_0)'$. It follows that under the null hypothesis,

$$\frac{\hat{\Delta}^2}{\boldsymbol{z}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}_0 MS_E} \sim F_{1,n-2},$$

where MS_E is the error mean square. Consequently,

$$P\left[\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0)^2}{\boldsymbol{z}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}_0 MS_E} \leq F_{\alpha,1,n-2}\right] = 1 - \alpha.$$

Note that

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix},$$

where n is the number of rows of $\boldsymbol{X}$. Hence,

$$\boldsymbol{z}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}_0 = \frac{\sum_{i=1}^n x_i^2 - 2x_0 \sum_{i=1}^n x_i + n x_0^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}.$$

The desired $(1 - \alpha)100\%$ confidence set consists of all the values of x_0 that satisfy the inequality

$$(\hat{\beta}_0 + \hat{\beta}_1 x_0)^2 \leq \boldsymbol{z}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}_0 MS_E F_{\alpha,1,n-2}.$$

11.6 **(a)** By Theorem 10.1,

$$\boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \sum_{i=1}^k \gamma_i W_i,$$

where the $\gamma_1, \gamma_2, \ldots, \gamma_k$ are the distinct nonzero eigenvalues of $\boldsymbol{V}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{1/2}$ with multiplicities $\nu_1, \nu_2, \ldots, \nu_k$, and the W_i's are mutually independent such that $W_i \sim \chi^2_{\nu_i}(\theta_i)$, where

$$\theta_i = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{V}^{-1/2}\boldsymbol{P}_i\boldsymbol{P}_i'\boldsymbol{V}^{-1/2}\boldsymbol{X}\boldsymbol{\beta},$$

and $\boldsymbol{P}_i$ is a matrix of order $n \times \nu_i$ whose columns are orthonormal eigenvectors of $\boldsymbol{V}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{1/2}$ corresponding to γ_i $(i = 1, 2, \ldots, k)$.

(b) The above applies with $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ replaced by $\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

11.7 This follows from writing $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ as $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}$ and noting that the square of its Euclidean norm is equal to $||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}||^2 + 2(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}) + ||\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}||^2$. The middle term is equal to zero. We conclude that

$$||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 \geq ||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}||^2.$$

Equality is achieved if and only if $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ since $||\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}||^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, and the right-hand side is zero if and only if $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ since $\boldsymbol{X}'\boldsymbol{X}$ is positive definite.

11.8

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_1 &= \boldsymbol{T}_{11}\boldsymbol{X}_1'\boldsymbol{y} + \boldsymbol{T}_{12}\boldsymbol{X}_2'\boldsymbol{y} \\
&= [\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{X}_2 \\
&\quad (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1]^{-1}\boldsymbol{X}_1'\boldsymbol{y} \\
&\quad -[\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1]^{-1} \\
&\quad \boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y} = \\
&\quad [\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1]^{-1}[\boldsymbol{X}_1'\boldsymbol{y} - \\
&\quad \boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}] \\
&= (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}.
\end{aligned}$$

11.9 **(a)** Let $n_1 = 2^k$, and $\boldsymbol{X}_1$ be the matrix of n_1 rows whose columns are associated with $x_1, x_2, \ldots, x_k$, and let the matrix $\boldsymbol{X}_2$ be the matrix of n_1 rows whose columns are associated with $x_ix_j (i < j)$ and $x_1^2, x_2^2, \ldots, x_k^2$. Then

$$\begin{aligned}
\mathrm{E}(\bar{y}_1) &= \boldsymbol{1}_{n_1}'\mathrm{E}(\boldsymbol{y}_1)/n_1 \\
&= \frac{1}{n_1}[\beta_0 n_1 + (n_1, n_1, \ldots, n_1)(\beta_{11}, \beta_{22}, \ldots, \beta_{kk})'] \\
&= \beta_0 + \sum_{i=1}^{k} \beta_{ii}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\bar{y}_0) &= \frac{1}{n_0}\boldsymbol{1}_{n_0}'(\boldsymbol{1}_{n_0}\beta_0) \\
&= \beta_0.
\end{aligned}$$

It follows that

$$\begin{aligned}\mathrm{E}(\bar{y}_1 - \bar{y}_0)^2 &= var(\bar{y}_1 - \bar{y}_0) + [\mathrm{E}(\bar{y}_1 - \bar{y}_0)]^2 \\ &= \sigma^2\left(\frac{1}{2^k} + \frac{1}{n_0}\right) + \left(\sum_{i=1}^{k} \beta_{ii}\right)^2 .\end{aligned}$$

(b)

$$MS_E = \frac{1}{N-k-1}\boldsymbol{y}'[\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y},$$

where $N = 2^k + n_0$, $\boldsymbol{X} = [\mathbf{1}_N : (\boldsymbol{X}_1' \ : \ \mathbf{0}')']$. Hence,

$$\mathrm{E}(MS_E) = \sigma^2 + \frac{1}{N-k-1}\mathrm{E}(\boldsymbol{y}')[\mathbf{1}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\mathrm{E}(\boldsymbol{y}).$$

We also have

$$\begin{aligned}\mathrm{E}(\boldsymbol{y}) &= \begin{bmatrix}\mathrm{E}(\boldsymbol{y}_1)\\ \mathrm{E}(\boldsymbol{y}_0)\end{bmatrix} \\ &= \begin{bmatrix}\beta_0\mathbf{1}_{n_1} + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 \\ \beta_0\mathbf{1}_{n_0}\end{bmatrix}.\end{aligned}$$

It can be seen that $\mathrm{E}(MS_E) \neq \sigma^2$. However, if $y_{01}, y_{02}, \ldots, y_{0n_0}$ are the center point observations, then the pure error sum of squares, SS_{PE}, is given by

$$\begin{aligned}SS_{PE} &= \sum_{i=1}^{n_0}(y_{0i} - \bar{y}_0)^2 \\ &= \sum_{i=1}^{n_0} y_{0i}^2 - n\bar{y}_0^2 \\ &= \boldsymbol{y}_0'\left(\boldsymbol{I}_{n_0} - \frac{1}{n_0}\boldsymbol{J}_{n_0}\right)\boldsymbol{y}_0.\end{aligned}$$

Thus, $MS_{PE} = \frac{1}{n_0-1}SS_{PE}$. Hence,

$$\begin{aligned}\mathrm{E}(MS_{PE}) &= \sigma^2 + \frac{\beta_0\mathbf{1}_{n_0}'\left(\boldsymbol{I}_{n_0} - \frac{1}{n_0}\boldsymbol{J}_{n_0}\right)\beta_0\mathbf{1}_{n_0}}{n_0 - 1} \\ &= \sigma^2.\end{aligned}$$

(c) $\frac{(\bar{y}_1-\bar{y}_0)^2}{(\frac{1}{n_1}+\frac{1}{n_0})MS_{PE}}$, where $n_1 = 2^k$, is a test statistic which, under the null hypothesis, has the F-distribution with 1 and $n_0 - 1$ degrees of freedom.

11.10 We have that

$$\hat{y}(\boldsymbol{x}_i) = \boldsymbol{z}'(\boldsymbol{x}_i)\hat{\boldsymbol{\beta}}_1, \quad i = 1, 2,$$

where $z'(x_i) = (1, x_i')$, $i = 1, 2$. Furthermore,

$$var(\hat{\beta}_1) = \sigma^2 (X'X)^{-1}$$
$$= \sigma^2 \begin{bmatrix} \frac{1}{n_1+n_0} & \mathbf{0} \\ \mathbf{0}' & \frac{1}{n_1} I_k \end{bmatrix}.$$

Hence,

$$var[\hat{y}(x_1) - \hat{y}(x_2)] = \sigma^2 [z'(x_1) - z'(x_2)] var(\hat{\beta}_1)[z(x_1) - z(x_2)]$$
$$= \sigma^2 (x_1' - x_2')(x_1 - x_2)/n_1$$
$$= \frac{\sigma^2}{n_1} \left[||x_1||^2 - 2||x_1|| \, ||x_2|| cos(\theta) + ||x_2||^2 \right].$$

11.11 The expected value of MS_E is

$$E(MS_E) = \frac{1}{n-p} \left[tr\{(I - X(X'X)^{-1}X')\Sigma\} + \beta' X'[I - X(X'X)^{-1}X']X\beta \right]$$
$$= \frac{1}{n-p} tr\{[I - X(X'X)^{-1}X']\Sigma\}$$
$$= \frac{1}{n-p} \{tr(\Sigma) - tr[X(X'X)^{-1}X'\Sigma]\}.$$

But,

$$tr[X(X'X)^{-1}X'\Sigma] = tr[X(X'X)^{-1}X'\Sigma X(X'X)^{-1}X'],$$

since $X(X'X)^{-1}X'$ is idempotent. The matrix $X(X'X)^{-1}X'\Sigma X(X'X)^{-1}X'$ is positive semidefinite since $u'X(X'X)^{-1}X'\Sigma X(X'X)^{-1}X'u \geq 0$, and is equal to zero if and only if $X(X'X)^{-1}X'u = \mathbf{0}$ and u does not have to be equal to a zero vector. Hence, $tr[X(X'X)^{-1}X'\Sigma] \geq 0$. It follows that $E[MS_E] \leq \frac{1}{n-p} tr(\Sigma)$.

11.12 Let $\hat{\theta} = \lambda' \hat{\beta}$, where

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y.$$

Note that $\hat{\theta}$ is unbiased for θ. Its variance, $var(\hat{\theta}) = \lambda'(X'\Sigma^{-1}X)^{-1}\lambda$ is known. Since $\hat{\theta}$ is normally distributed, then

$$\frac{\hat{\theta} - \theta}{[\lambda'(X'\Sigma^{-1}X)^{-1}\lambda]^{1/2}} \sim N(0, 1).$$

Hence, a $(1 - \alpha)100\%$ confidence interval on θ is $\hat{\theta} \pm [\lambda'(X'\Sigma^{-1}X)^{-1}\lambda]^{1/2} z_{\frac{\alpha}{2}}$.

11.13 (i) $\hat{y}(x_0) - y(x_0) \sim N(0, \sigma^2 z'(x_0)(X'X)^{-1}z(x_0) + \sigma^2)$, where $z(x_0)$ is such that $\hat{y}(x_0) = z'(x_0)\hat{\beta}$. Hence,

$$\frac{\hat{y}(x_0) - y(x_0)}{\{\sigma^2[1 + z'(x_0)(X'X)^{-1}z(x_0)]\}^{1/2}} \sim N(0, 1).$$

We also have that $\frac{(n-p)MS_E}{\sigma^2} \sim \chi^2_{n-p}$. Hence,

$$\frac{\hat{y}(\boldsymbol{x}_0) - y(\boldsymbol{x}_0)}{\{MS_E[1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)]\}^{1/2}} \sim t_{n-p}.$$

Consequently, the given probability can be computed using

$$P\left[t_{n-p} > \frac{1}{[1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)]^{1/2}}\right].$$

(ii) (a) $d = var[\hat{y}(\boldsymbol{x}_0) - y(\boldsymbol{x}_0)] = \sigma^2[1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)]$.

(b) $\hat{d} = MS_E[1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)]$.

(c) $$P\left[\hat{d} > 2\sigma^2\right] = P\left\{\frac{MS_E[1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)]}{\sigma^2} > 2\right\}.$$

Since, $\frac{(n-p)MS_E}{\sigma^2} \sim \chi^2_{n-p}$, then

$$P[\hat{d} > 2\sigma^2] = P\left\{\chi^2_{n-p} > \frac{2(n-p)}{1 + \boldsymbol{z}'(\boldsymbol{x}_0)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{z}(\boldsymbol{x}_0)}\right\}.$$

CHAPTER 12

12.1 Given that model(12.71) is of full column rank, the vector $\boldsymbol{\delta}$ in (12.71) is an estimable linear function of $\boldsymbol{\beta}$. Since the r rows of the matrix $\boldsymbol{P}_1'$ are linearly independent, we conclude that there are only $r(\boldsymbol{X})$ linearly independent estimable functions.

12.2 We have that

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{P}_1\boldsymbol{\Lambda}\boldsymbol{P}_1'.\end{aligned}$$

It is then clear that the rows of $\boldsymbol{X}$ are linear combinations of the rows of $\boldsymbol{P}_1'$.

12.3 Since $\boldsymbol{Q}'\boldsymbol{\beta}$ is estimable, there exists a matrix $\boldsymbol{T}'$ such that $\boldsymbol{Q}' = \boldsymbol{T}'\boldsymbol{X}$. Hence, $\boldsymbol{Q}'\boldsymbol{\beta}^{\mathbf{0}} = \boldsymbol{T}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ and the matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is invariant to the choice of $(\boldsymbol{X}'\boldsymbol{X})^{-}$.

12.4 Let $\boldsymbol{\lambda}'\boldsymbol{y}$ be another unbiased estimator of $\boldsymbol{a}'\boldsymbol{\beta}$ (other that $\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}$). Then, $\boldsymbol{\lambda}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{\beta}$. Hence, $\boldsymbol{a}' = \boldsymbol{\lambda}'\boldsymbol{X}$. Also, $var(\boldsymbol{\lambda}'\boldsymbol{y}) = \boldsymbol{\lambda}'\boldsymbol{\lambda}\sigma^2$. Furthermore,

$$\begin{aligned}var(\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}) &= \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{a}\sigma^2 \\ &= \boldsymbol{\lambda}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\lambda}\sigma^2 \\ &\le \boldsymbol{\lambda}'\boldsymbol{\lambda}\sigma^2.\end{aligned}$$

The inequality is true since the matrix $\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is a symmetric idempotent matrix, hence positive semidefinite. We therefore conclude that $var(\boldsymbol{a}'\boldsymbol{\beta}^{\mathbf{0}}) \leq var(\boldsymbol{\lambda}'\boldsymbol{y})$.

12.5 **(a)** $\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$ is estimable since $\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}$ belongs to the row space of $\boldsymbol{X}$. Vice versa, if $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}' = \boldsymbol{t}'\boldsymbol{X} = \boldsymbol{t}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}$, where $\boldsymbol{\ell}' = \boldsymbol{t}'\boldsymbol{X}$.

(b) This is obvious since all coefficients are estimable and $r[(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}] = r(\boldsymbol{X}'\boldsymbol{X}) = r(\boldsymbol{X})$.

12.6 **(a)** Solutions to the normal equations can be obtained from the SAS output under parameter estimates. For the parameters $\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2$, and β_3 we find 15, 9, -2, -1, 0, -8.5, -9.5, and 0, respectively.

(b) Let $\mu_{ij} = \mu + \alpha_i + \beta_j$. Note that $\beta_1 - \beta_2 = \mu_{11} - \mu_{12}$ is estimable as well as $\beta_1 - \beta_3 = \mu_{11} - \mu_{13}$ since the cell means, $\mu_{11}, \mu_{12}, \mu_{13}$ are estimable because their corresponding cells are nonempty. Furthermore, the F-test value in (12.48) is equal to (9.91) which is significant at a level ≥ 0.0126.

(c) The power value, as described in Section 12.10.3, is equal to 0.845. This can be seen from the SAS output using PROC IML.

(d) $\mu + \frac{1}{4}\sum_{i=1}^{4} \alpha_i + \beta_1 = \frac{1}{4}\sum_{i=1}^{4} \mu_{i1}$ is estimable since the cell means are estimable because their corresponding cells are nonempty. Using the values given in part (a) we find that the BLUE of $\mu + \frac{1}{4}\sum_{i=1}^{4} \alpha_i + \beta_1$ is equal to 8.0.

12.7 We have that $\mu_{21} = \mu + \alpha_2 + \beta_1 = \mu + \alpha_2 + \beta_3 + (\beta_1 - \beta_3) = \mu_{23} + (\beta_1 - \beta_3)$, which is estimable since cell (2,3) is nonempty, and $\beta_1 - \beta_3$ is estimable given, for example, the cells in row 1 are nonempty.
$\mu_{22} = \mu + \alpha_2 + \beta_2 = \mu + \alpha_2 + \beta_3 + (\beta_2 - \beta_3) = \mu_{23} + (\beta_2 - \beta_3)$. Both μ_{23} and $(\beta_2 - \beta_3)$ are estimable.
$\mu_{32} = \mu + \alpha_3 + \beta_2 = \mu + \alpha_3 + \beta_3 + (\beta_2 - \beta_3)$, which are estimable.

12.8 ϕ is estimable since $\phi = \sum_{i=1}^{4} \lambda_i\mu_{i3} = \mu\sum_{i=1}^{4} \lambda_i + \sum_{i=1}^{4} \lambda_i\alpha_i + \beta_3\sum_{i=1}^{4} \lambda_i$, and the four cell means for column 3 are all estimable.
The BLUE can be obtained by using solutions from the normal equations for the α_i's.

12.9 **(a)** $R(\mu) = N\bar{y}_{..}^2 = 12(10.5)^2 = 1323$.

(b) $R(\alpha|\mu) = R(\mu, \alpha) - R(\mu) = \sum_{i=1}^{4} n_i\bar{y}_{i.}^2 - 1323 = 3(18^2 + 7^2 + 8^2 + 9^2) - 1323 = 231$.

(c) $R(\mu, \alpha) = 1554$.

(d) $R(\alpha|\mu, \beta) = 231$.

(e) $R(\beta|\mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha) = 1772 - 1554 = 218$.

12.10 **(a)** The number of basic linearly independent estimable functions is equal to the rank of the matrix $\boldsymbol{X}$ in the model. This rank is equal to 6 since there are 8 columns in $\boldsymbol{X}$ with two linearly independent relationships among the columns (the α_i columns sum to $\mathbf{1}_8$ and the β_j columns sum to $\mathbf{1}_8$). The E option in SAS's model statement prints the general form of estimable functions, where the number of distinct L_i values is the same as the number of basic estimable linear functions for the model. It can be seen that $\mu + \alpha_1 + \beta_2 = \mu + \alpha_1 + \beta_3 + \beta_2 - \beta_3$, $\alpha_1 - \alpha_3$, $\alpha_2 - \alpha_3$, $\beta_1 - \beta_2$, $\beta_2 - \beta_3$, $\beta_3 - \beta_4$ form a set of basic estimable linear functions.

(b) $\mu_{12} = \mu + \alpha_1 + \beta_2 = \mu + \alpha_1 + \beta_3 + (\beta_2 - \beta_3) = \mu_{13} + (\beta_2 - \beta_3)$, which is estimable.

(c) $\phi = \frac{1}{4}\sum_{j=1}^{4}(\mu + \alpha_i + \beta_j)$ is estimable for all i since μ_{12} is estimable, so are $\mu_{21} = \mu + \alpha_2 + \beta_3 + (\beta_1 - \beta_3)$, $\mu_{22} = \mu + \alpha_2 + \beta_2 = \mu + \alpha_2 + \beta_3 + (\beta_2 - \beta_3)$, $\mu_{34} = \mu + \alpha_3 + \beta_4 = \mu + \alpha_3 + \beta_3 + (\beta_4 - \beta_3)$. The BLUE of ϕ can be obtained by replacing the parameters in ϕ with their corresponding solutions from the SAS output under Parameter Estimate. These solutions are $\mu^0 = 28.4$, $\alpha_1^0 = 0.2$, $\alpha_2^0 = -16.0$, $\alpha_3^0 = 0.0$, $\beta_1^0 = -12.0$, $\beta_2^0 = -13.4$, $\beta_3^0 = -7.8$, $\beta_4^0 = 0.0$.

CHAPTER 13

13.1 **(a)** (i:k)(j)

(b)
$$y_{ijk} = \mu + \alpha_i + \pi_{i(k)} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{i(jk)}.$$

(c) The following is the ANOVA table:

Source	DF	SS	MS	E(MS)
A	$p-1$	SS_A	MS_A	$\frac{qn}{p-1}\sum_{i=1}^{p}\alpha_i^2 + q\sigma^2_{\pi(\alpha)} + \sigma^2_\epsilon$
$P(A)$	$p(n-1)$	$SS_{P(A)}$	$MS_{P(A)}$	$q\sigma^2_{\pi(\alpha)} + \sigma^2_\epsilon$
B	$q-1$	SS_B	MS_B	$\frac{pn}{q-1}\sum_{j=1}^{q}\beta_j^2 + \sigma^2_\epsilon$
$A*B$	$(p-1)(q-1)$	SS_{AB}	MS_{AB}	$\frac{n}{(p-1)(q-1)}\sum_{i=1}^{p}\sum_{j=1}^{q}(\alpha\beta)^2_{(ij)} + \sigma^2_\epsilon$
Error	$p(q-1(n-1)$	SS_E	MS_E	σ^2_ϵ

13.2 **(a)** $[(i : k)(j)] : l$

(b)
$$y_{ijkl} = \mu + \alpha_{(i)} + \gamma_{i(k)} + \beta_{(j)} + (\alpha\beta)_{(ij)} + (\beta\gamma)_{i(jk)} + \epsilon_{ijk(l)}.$$

(c) The ANOVA table is

Source	DF	SS	MS	E(MS)
A	1	73.5	73.5	$12\sum_{i=1}^{2}\alpha_i^2 + 4\sigma^2_{\gamma(\alpha)} + 2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon$
$C(A)$	4	672.83	168.208	$4\sigma^2_{\gamma(\alpha)} + 2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon$
B	1	962.667	962.667	$12\sum_{j=1}^{2}\beta_j^2 + 2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon$
$A*B$	1	4.1667	4.1667	$6\sum_{i=1}^{2}\sum_{j=1}^{2}(\alpha\beta)^2_{ij} + 2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon$
$B*C(A)$	4	258.167	64.542	$2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon$
Error	12	60.0	5.0	σ^2_ϵ

(d) $H_0 : \alpha_i = 0, F = \frac{MS_A}{MS_{C(A)}} = 0.44, p - value = 0.5447.$

$H_0 : \sigma^2_{\gamma(\alpha)} = 0, F = \frac{MS_{C(A)}}{MS_{B*C(A)}} = 2.61, p - value = 0.1880.$

$H_0 : \beta_j = 0, F = \frac{MS_B}{MS_{B*C(A)}} = 14.92, p - value = 0.0181.$

$H_0 : (\alpha\beta)_{ij} = 0, F = \frac{MS_{A*B}}{MS_{B*C(A)}} = 0.06, p - value = 0.812.$

$H_0 : \sigma^2_{\beta\gamma(\alpha)} = 0, F = \frac{MS_{B*C(A)}}{MS_E} = 12.91, p - value = 0.0003.$

13.3 An unbiased estimate of $\sigma^2_{\beta\gamma(\alpha)}$ is given by

$$\hat{\sigma}^2_{\beta\gamma(\alpha)} = \frac{MS_{B*C(A)} - MS_E}{2} = 29.771.$$

Thus, we approximately have

$$\frac{\nu \ \hat{\sigma}^2_{\beta\gamma(\alpha)}}{\sigma^2_{\beta\gamma(\alpha)}} \sim \chi^2_\nu,$$

where

$$\begin{aligned}\nu &= \frac{(MS_{B*C(A)} - MS_E)^2}{\frac{(MS_{B*C(A)})^2}{4} + \frac{(-MS_E)^2}{12}} \\ &= 3.397.\end{aligned}$$

Hence, we approximately have

$$\frac{3.397 \ \hat{\sigma}^2_{\beta\gamma(\alpha)}}{\sigma^2_{\beta\gamma(\alpha)}} \sim \chi^2_3.$$

We can then write

$$P\left[\chi^2_{0.975,3} < \frac{3.397 \ \hat{\sigma}^2_{\beta\gamma(\alpha)}}{\sigma^2_{\beta\gamma(\alpha)}} < \chi^2_{0.025,3}\right] \approx 0.95.$$

This produces an approximate 95% confidence interval on $\sigma^2_{\beta\gamma(\alpha)}$ given by

$$10.82 < \sigma^2_{\beta\gamma(\alpha)} < 459.69.$$

13.4 We have by Theorem 13.1,

$$\frac{4MS_{B*C(A)}}{2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon} \sim \chi^2_4,$$

and $\frac{12MS_E}{\sigma_\epsilon^2} \sim \chi^2_{12}$. Hence, with probability 0.95,

$$\frac{MS_{B*C(A)}}{MS_E}F_{0.975,12,4} < \frac{2\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_\epsilon}{\sigma^2_\epsilon} < \frac{MS_{B*C(A)}}{MS_E}F_{0.025,12,4},$$

which can be written as

$$\frac{64.542}{5.0}\frac{1}{F_{0.025,4,12}} < 1 + \frac{2\sigma^2_{\beta\gamma(\alpha)}}{\sigma^2_\epsilon} < \frac{64.542}{5.0} 8.75.$$

This gives

$$2.065 < \frac{\sigma^2_\epsilon + \sigma^2_{\beta\gamma(\alpha)}}{\sigma^2_\epsilon} < 56.975.$$

Since

$$\frac{\sigma^2_{\beta\gamma(\alpha)}}{\sigma^2_\epsilon + \sigma^2_{\beta\gamma(\alpha)}} = 1 - \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + \sigma^2_{\beta\gamma(\alpha)}},$$

we conclude that the desired interval is

$$0.516 < \frac{\sigma^2_{\beta\gamma(\alpha)}}{\sigma^2_\epsilon + \sigma^2_{\beta\gamma(\alpha)}} < 0.9824.$$

13.5 **(a)** $(i : j : k : l)$

(b) The expected mean squares are given in the following ANOVA table:

Source	DF	E(MS)
A	$a-1$	$\frac{bcn}{a-1}\sum_{i=1}^a \alpha_i^2 + cn\sigma^2_{\beta(\alpha)} + n\sigma^2_{\gamma(\alpha\beta)} + \sigma^2_\epsilon$
$B(A)$	$a(b-1)$	$+ cn\sigma^2_{\beta(\alpha)} + n\sigma^2_{\gamma(\alpha\beta)} + \sigma^2_\epsilon$
$C(AB)$	$ab(c-1)$	$n\sigma^2_{\gamma(\alpha\beta)} + \sigma^2_\epsilon$
Error	$abc(n-1)$	σ^2_ϵ

(c) We have,

$$\frac{SS_A}{cn\sigma^2_{\beta(\alpha)} + n\sigma^2_{\gamma(\alpha\beta)} + \sigma^2_\epsilon} \sim \chi^2_{a-1}(\lambda),$$

$$\frac{SS_{B(A)}}{cn\sigma^2_{\beta(\alpha)} + n\sigma^2_{\gamma(\alpha\beta)} + \sigma^2_\epsilon} \sim \chi^2_{a(b-1)}.$$

where

$$\lambda = \frac{bcn \sum_{i=1}^{a} \alpha_i^2}{cn\sigma_{\beta(\alpha)}^2 + n\sigma_{\gamma(\alpha\beta)}^2 + \sigma_\epsilon^2}.$$

The probability of making a Type II error is

$$P\left[\frac{MS_A}{MS_{B(A)}} < F_{0.05,a-1,a(b-1)} | H_a\right]$$
$$= P[F'_{a-1,a(b-1)}(\lambda) < F_{0.05,a-1,a(b-1)}].$$

(d) $\hat{\sigma}_{\beta(\alpha)}^2 = \frac{1}{cn}[MS_{B(A)} - MS_{C(AB)}]$.

$$P[\hat{\sigma}_{\beta(\alpha)}^2 < 0] = P[MS_{B(A)} < MS_{C(AB)}].$$

We also have

$$\frac{SS_{B(A)}}{cn\sigma_{\beta(\alpha)}^2 + n\sigma_{\gamma(\alpha\beta)}^2 + \sigma_\epsilon^2} \sim \chi_{a(b-1)}^2$$
$$\frac{SS_{C(AB)}}{n\sigma_{\gamma(\alpha\beta)}^2 + \sigma_\epsilon^2} \sim \chi_{ab(c-1)}^2.$$

We conclude that

$$P[\hat{\sigma}_{\beta(\alpha)}^2 < 0] = P\left[F_{a(b-1),ab(c-1)} < \frac{n\sigma_{\gamma(\alpha\beta)}^2 + \sigma_\epsilon^2}{cn\sigma_{\beta(\alpha)}^2 + n\sigma_{\gamma(\alpha\beta)}^2 + \sigma_\epsilon^2}\right].$$

13.6 The test statistic is $F = \frac{MS_\alpha + MS_E}{MS_{\alpha\beta} + MS_{\alpha\gamma}} = 2.546$. The numerator and denominator degrees of freedom are approximately given by ν_1 and ν_2, respectively.

$$\nu_1 = \frac{(MS_\alpha + MS_E)^2}{\frac{(MS_\alpha)^2}{4} + \frac{(MS_E)^2}{24}}$$
$$= 4.68 \approx 5.$$

$$\nu_2 = \frac{(MS_{\alpha\beta} + MS_{\alpha\gamma})^2}{\frac{(MS_{\alpha\beta})^2}{12} + \frac{(MS_{\alpha\gamma})^2}{8}}$$
$$= 19.57 \approx 20.$$

The 10% critical value is approximately equal to $F_{0.10,5,20} = 2.16$. Hence, the test is significant at the approximate 10% level.

13.7 **(a)** We have that $\frac{MS_A/(n\sigma_\alpha^2+\sigma_\epsilon^2)}{MS_E/\sigma_\epsilon^2} \sim F_{a-1,a(n-1)}$. Hence, with a probability equal to $1-\alpha$, we have

$$F_{1-\frac{\alpha}{2},a-1,a(n-1)} < \frac{\sigma_\epsilon^2}{n\sigma_\alpha^2+\sigma_\epsilon^2}\frac{MS_A}{MS_E} < F_{\frac{\alpha}{2},a-1,a(n-1)}.$$

From this we get the interval

$$\frac{1}{n}\left[\frac{MS_A}{MS_E}\frac{1}{F_{\frac{\alpha}{2},a-1,a(n-1)}} - 1\right] < \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} < \frac{1}{n}\left[\frac{MS_A}{MS_E}\frac{1}{F_{1-\frac{\alpha}{2},a-1,a(n-1)}} - 1\right].$$

(b) If we denote the lower and upper end points of the interval in (a) by L_1 and L_2, respectively, the desired interval here is given by

$$\frac{1}{1+L_2} < \frac{\sigma_\epsilon^2}{\sigma_\alpha^2+\sigma_\epsilon^2} < \frac{1}{1+L_1}.$$

13.8 **(a)** $i : [(j)(k : l)]$

(b) The ANOVA table is

(c) We have

$$F = \frac{MS_{D(AC)}/[3\sigma_{\delta(\alpha\gamma)}^2+\sigma_\epsilon^2]}{MS_E/\sigma_\epsilon^2} \sim F_{24,48}$$

Source	DF	MS	E(MS)
A	2	MS_A	$18\sum_{i=1}^3 \alpha_i^2 + 12\sigma_{\beta(\alpha)}^2 + 9\,\sigma_{\gamma(\alpha)}^2 + 3\sigma_{\delta(\alpha\gamma)}^2 + 3\sigma_{\beta\gamma(\alpha)}^2 + \sigma_\epsilon^2$
$B(A)$	6	$MS_{B(A)}$	$12\sigma_{\beta(\alpha)}^2 + 3\sigma_{\beta\gamma(\alpha)}^2 + \sigma_\epsilon^2$
$C(A)$	9	$MS_{C(A)}$	$9\sigma_{\gamma(\alpha)}^2 + 3\,\sigma_{\delta(\alpha\gamma)}^2 + 3\,\sigma_{\beta\gamma(\alpha)}^2 + \sigma_\epsilon^2$
$D(AC)$	24	$MS_{D(AC)}$	$3\sigma_{\delta(\alpha\gamma)}^2 + \sigma_\epsilon^2$
$B * C(A)$	18	$MS_{B*C(A)}$	$3\sigma_{\beta\gamma(\alpha)}^2 + \sigma_\epsilon^2$
$Error$	48	MS_E	σ_ϵ^2

Under $H_0 : \sigma_{\delta(\alpha\gamma)}^2 = 2\sigma_\epsilon^2$,

$$F = \frac{MS_{D(AC)}/7\sigma_\epsilon^2}{MS_E/\sigma_\epsilon^2} = \frac{1}{7}\frac{MS_{D(AC)}}{MS_E} \sim F_{24,48}$$

We can reject H_0 at the α-level if $F > F_{\frac{\alpha}{2},24,48}$ or $F < F_{1-\frac{\alpha}{2},24,48}$.

13.9 **(a)** We have

$$\frac{SS_{B(A)}}{\theta_1} \sim \chi_6^2$$
$$\frac{SS_{B*C(A)}}{\theta_2} \sim \chi_{18}^2$$

Let $\alpha^* = 1 - (1 - \alpha)^{1/2}$. We then get

$$\frac{SS_{B(A)}}{\chi_{\frac{\alpha*}{2},6}} < \quad \theta_1 \quad < \frac{SS_{B(A)}}{\chi_{1-\frac{\alpha^*}{2},6}}$$
$$\frac{SS_{B*C(A)}}{\chi_{\frac{\alpha^*}{2},18}} < \quad \theta_2 \quad < \frac{SS_{B*C(A)}}{\chi_{1-\frac{\alpha^*}{2},18}}$$

with a joint probability of $1 - \alpha$. If the end points of the first and second intervals are denoted by u_1, u_2 and v_1, v_2, respectively, then we get

$$P[\theta \in (u_1, u_2) \times (v_1, v_2)] = (1 - \alpha^*)^2 = 1 - \alpha.$$

(b) We have that

$$\hat{\sigma}^2_{\beta(\alpha)} = \frac{MS_{B(A)} - MS_{B*C(A)}}{12}.$$

Hence,

$$\begin{aligned}
P\left(\hat{\sigma}^2_{\beta(\alpha)} < 0\right) &= P\left[\frac{MS_{B(A)}}{MS_{B*C(A)}} < 1\right] \\
&= P\left[\frac{MS_{B(A)}/\left(12\sigma^2_{\beta(\alpha)} + 3\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_{\epsilon}\right)}{MS_{B*C(A)}/\left(3\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_{\epsilon}\right)} < \frac{3\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_{\epsilon}}{12\sigma^2_{\beta(\alpha)} + 3\sigma^2_{\beta\gamma(\alpha)} + \sigma^2_{\epsilon}}\right] \\
&= P\left[F_{6,18} < \frac{1}{12\Delta + 1}\right].
\end{aligned}$$

13.10 **(a)** $[(i)(j) : k] : l$

(b) The admissible means and corresponding components are given in Table 1. The model is then of the form

$$y_{ijkl} = \mu + \alpha_{(i)} + \beta_{(j)} + (\alpha\beta)_{(ij)} + \delta_{ij(k)} + \epsilon_{ijk(l)}.$$

The corresponding ANOVA table is given Table 2.

Table 1

Admissible mean	Component
y	y
$y_{(i)}$	$y_{(i)} - y$
$y_{(j)}$	$y_{(j)} - y$
$y_{(ij)}$	$y_{(ij)} - y_{(i)} - y_{(j)} + y$
$y_{ij(k)}$	$y_{ij(k)} - y_{(ij)}$
$y_{ijk(l)}$	$y_{ijk(l)} - y_{ij(k)}$

Table 2

Source	DF	$E(MS)$
A	2	$16\sum_{i=1}^{3}\alpha_i^2 + 8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2$
B	3	$24\sigma_\beta^2 + 8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2$
$A*B$	6	$8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2$
$C(AB)$	36	$2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2$
Error	48	σ_ϵ^2

(c) $\hat{\sigma}_{\alpha\beta}^2 = \frac{MS_{A*B} - MS_{C(AB)}}{8}$. We also have

$$\frac{SS_{AB}}{8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2} \sim \chi_6^2$$

$$\frac{SS_{C(AB)}}{2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2} \sim \chi_{36}^2$$

Hence,

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{1}{8}\left[\frac{8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2}{6}\chi_6^2 - \frac{2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2}{36}\chi_{36}^2\right].$$

It follows that

$$var(\hat{\sigma}_{\alpha\beta}^2) = \frac{1}{64}\left[\left(\frac{8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2}{6}\right)^2 \times (12) + \left(\frac{2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2}{36}\right)^2 \times (72)\right].$$

(d) We have that

$$\bar{y}_{i\ldots} = \mu + \alpha_i + \bar{\beta}_{.} + \overline{(\alpha\beta)}_{i.} + \bar{\delta}_{i..} + \bar{\epsilon}_{i\ldots}.$$

Hence,

$$var(\bar{y}_{1\ldots} - \bar{y}_{2\ldots}) = \frac{2}{32}\left[8\sigma_{\alpha\beta}^2 + 2\sigma_{\delta(\alpha\beta)}^2 + \sigma_\epsilon^2\right].$$

It follows that a $(1-\alpha)100\%$ confidence interval on $\alpha_{(1)} - \alpha_{(2)}$ is given by

$$\bar{y}_{1\ldots} - \bar{y}_{2\ldots} \pm \left(\frac{2MS_{A*B}}{32}\right)^{1/2} t_{\frac{\alpha}{2},6}.$$

13.11 **(a)** We have that $\sum_{i=0}^{q+1} \boldsymbol{P}_i = \boldsymbol{I}_N$. Hence,

$$\begin{aligned} \boldsymbol{X} &= \left(\sum_{i=0}^{q+1} \boldsymbol{P}_i\right)\boldsymbol{X} \\ &= \sum_{i=0}^{q+1} (\boldsymbol{P}_i\boldsymbol{X}). \end{aligned}$$

But, $\boldsymbol{P}_i\boldsymbol{X} = \boldsymbol{0}$ for $i = q-p+1, \ldots, q+1$, that is, if the effect is random. This is true because $\boldsymbol{P}_i\boldsymbol{A}_j = \boldsymbol{0}$ for $j = 0, 1, \ldots, q-p$ by formula (13.19) since a set of subscripts, η_i, associated with a random effect cannot be a subset of η_j for a fixed effect. Thus, $\boldsymbol{P}_i\boldsymbol{U}_j\boldsymbol{U}_j' = \boldsymbol{0}$ which implies, $\boldsymbol{P}_i\boldsymbol{U}_j\boldsymbol{U}_j'\boldsymbol{U}_j = \boldsymbol{0}$. Consequently, $\boldsymbol{P}_i\boldsymbol{U}_j(b_i\boldsymbol{I}_{c_j}) = \boldsymbol{0}$ by (13.8). We then conclude that $\boldsymbol{P}_i\boldsymbol{U}_j = \boldsymbol{0}$ for $i = q-p+1, \ldots, q+1, j = 0, 1, \ldots, q-p$. This results in $\boldsymbol{P}_i\boldsymbol{X} = \boldsymbol{P}_i[\boldsymbol{U}_0 : \boldsymbol{U}_1 : \ldots : \boldsymbol{U}_{q-p}] = \boldsymbol{0}$. Consequently, $\boldsymbol{X} = \sum_{i=0}^{q-p} \boldsymbol{P}_i\boldsymbol{X}$.

(b) Suppose that there exit constants, $a_0, a_1, a_2, \ldots, a_{q+1}$, not all equal to zero such that $\sum_{i=0}^{q+1} a_i\boldsymbol{P}_i = \boldsymbol{0}$. Multiplying both sides by $\boldsymbol{P}_j$, we get, $a_i\boldsymbol{P}_j = \boldsymbol{0}$. This implies that $a_j = 0$ for all j. We conclude that the $\boldsymbol{P}_i$'s are linearly independent.

(c) We have that

$$\begin{aligned} r\left(\sum_{i=0}^{q-p} \boldsymbol{P}_i\right) &= r\left(\sum_{i=0}^{q-p} \boldsymbol{P}_i^2\right) \\ &= r\{[\boldsymbol{P}_0 : \boldsymbol{P}_1 : \ldots : \boldsymbol{P}_{q-p}][\boldsymbol{P}_0 : \boldsymbol{P}_1 : \ldots : \boldsymbol{P}_{q-p}]'\} \\ &= r[\boldsymbol{P}_0 : \boldsymbol{P}_1 : \ldots : \boldsymbol{P}_{q-p}] \\ &= \sum_{i=0}^{q-p} m_i, \end{aligned}$$

since by part (b) the columns of the $\boldsymbol{P}_i$'s are linearly independent and $r(\boldsymbol{P}_i) = m_i$.

CHAPTER 14

14.1 If $\boldsymbol{B}\boldsymbol{Y}' = \boldsymbol{1}_n' \otimes \boldsymbol{c}$, then

$$\begin{aligned} \boldsymbol{B}\boldsymbol{G} &= \boldsymbol{B}\boldsymbol{Y}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right) \\ &= (\boldsymbol{1}_n' \otimes \boldsymbol{c})\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right). \end{aligned}$$

The ith row of $\mathbf{1}'_n \otimes \boldsymbol{c}$ is of the form $c_i\mathbf{1}'_n$, where c_i is the ith element of $\boldsymbol{c}$ ($i = 1, 2, \ldots, m$). Since $c_i\mathbf{1}'_n(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n) = \mathbf{0}$ for $i = 1, 2, \ldots, m$, then $\boldsymbol{BG} = \mathbf{0}$, that is, m linearly independent relationships exit among the rows of $\boldsymbol{G}$. Thus, the rank of $\boldsymbol{G}$ is $r - m$, that is, the rank of $\boldsymbol{GG}'$ is $r - m$. Hence, $\boldsymbol{GG}'$ has a zero eigenvalue of multiplicity m.

Vice versa, if $\boldsymbol{GG}'$ has a zero eigenvalue of multiplicity m, then

$$\boldsymbol{GG}' = (\boldsymbol{P}_1\ \boldsymbol{P}_2)diag(\mathbf{0}, \boldsymbol{\Lambda})(\boldsymbol{P}_1\ \boldsymbol{P}_2)'.$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ contain columns of eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of nonzero eigenvalues of $\boldsymbol{GG}'$. This implies that $\boldsymbol{GG}'\boldsymbol{P}_1 = \mathbf{0}$, which gives $\boldsymbol{P}'_1\boldsymbol{G} = \mathbf{0}$, that is, $\boldsymbol{P}'_1\boldsymbol{Y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n) = \mathbf{0}$. We thus have

$$\boldsymbol{P}'_1\boldsymbol{Y}' = \frac{1}{n}\boldsymbol{P}'_1\boldsymbol{Y}'\boldsymbol{J}_n.$$

This relationship is of the form given in (14.5) with $\boldsymbol{B} = \boldsymbol{P}'_1$ and $\boldsymbol{c} = \frac{1}{n}\boldsymbol{P}'_1\boldsymbol{Y}'\mathbf{1}_n$. We thus have m linearly independent relationships of the type given in (14.5), with $\frac{1}{n}\boldsymbol{P}'_1\boldsymbol{Y}'\boldsymbol{J}_n$ having identical columns, exist among the responses.

14.2 We have that $\boldsymbol{GG}' = (\boldsymbol{P}_1\ \boldsymbol{P}_2)diag(\mathbf{0}, \boldsymbol{\Lambda})(\boldsymbol{P}_1\ \boldsymbol{P}_2)'$. Thus, $\boldsymbol{GG}'\boldsymbol{P}_1 = \mathbf{0}$, which implies $\boldsymbol{G}'\boldsymbol{P}_1 = \mathbf{0}$, that is, $(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{YP}_1 = \mathbf{0}$. Let $\boldsymbol{p}_{j1}$ be the jth column of $\boldsymbol{P}_1$. Then, $(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{Yp}_{j1} = \mathbf{0}$ for $j = 1, 2, \ldots, m$. We then have

$$\boldsymbol{Yp}_{j1} = \frac{1}{n}\mathbf{1}_n(\mathbf{1}'_n\boldsymbol{Yp}_{j1}), \quad j = 1, 2, \ldots, m,$$

Therefore, the elements of $\boldsymbol{p}_{j1}$ define a linear relationship among the r responses of the form

$$\sum_{i=1}^{r} y_{ui}p_{ij1} = \frac{1}{n}(\mathbf{1}'_n\boldsymbol{Yp}_{j1}), \quad j = 1, 2, \ldots, m;\ u = 1, 2, \ldots, n,$$

where y_{ui} is the (u, i)th element of $\boldsymbol{Y}$, p_{ij1} is the ith element of $\boldsymbol{p}_{j1}$. Thus, the m eigenvectors corresponding to a zero eigenvalue of $\boldsymbol{GG}'$ define m linearly independent relationships among the responses.

14.3 $\boldsymbol{G} = \boldsymbol{Y}'(\boldsymbol{I}_{13} - \frac{1}{13}\boldsymbol{J}_{13})$, $\delta = 0.005$, $\sigma^2_{re} = \delta^2/3 = 0.000008$. $E(\lambda) = (n-1)\sigma^2_{re} = 0.000096$. We also have

$$\sigma^2_\lambda \leq \left[\frac{9nr}{5} + nr(nr-1) - (n-1)^2\right]\sigma^4_{re} = 9(10)^{-8}.$$

Hence, we approximately have $\sigma_\lambda \leq 0.0003$. The eigenvalues of $\boldsymbol{GG}'$ are $0.0224814, 0.2628278, 6.4675216$. The distance of 0.0224814 from $E(\lambda)$ in units of σ_λ is

$$\frac{0.0224814 - 0.000096}{0.0003} = 74.6.$$

We conclude that $\boldsymbol{GG}'$ does not have zero eigenvalues and, therefore, no linear relationships exist among the three responses.

14.4 $\boldsymbol{G} = \boldsymbol{Y}'(\boldsymbol{I}_8 - \frac{1}{8}\boldsymbol{J}_8)$. The eigenvalues of $\boldsymbol{GG}'$ are $0.0670614, 3.0839231, 14.855072, 8828.7577$. $\delta = 0.05$, $\sigma^2_{re} = \delta^2/3 = 0.00083$. $E(\lambda) = (n-1)\sigma^2_{re} = 0.0058$.

$$\sigma^2_\lambda \leq \left[\frac{9nr}{5} + nr(nr-1) - (n-1)^2\right]\sigma^4_{re}.$$

Hence, $\sigma_\lambda \leq 0.02625$. The distance of 0.0670614 from 0.0058 in units of 0.02625 is $\frac{0.0670614-0.0058}{0.02625} = 2.334$. Therefore, 0.06760614 does not correspond to a zero eigenvalue. We conclude that no linear relationships exist among the responses.

14.5 $\boldsymbol{X} = \boldsymbol{I}_r \otimes \boldsymbol{X}_0$. Then,

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[(\boldsymbol{I}_r \otimes \boldsymbol{X}_0')(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_r \otimes \boldsymbol{X}_0)\right]^{-1}(\boldsymbol{I}_r \otimes \boldsymbol{X}_0')(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \\
&= \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_0'\boldsymbol{X}_0\right)^{-1}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_0')\boldsymbol{y} \\
&= \left[\boldsymbol{I}_r \otimes (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\right]\boldsymbol{y}.
\end{aligned}$$

14.6

$$\frac{\partial}{\partial \boldsymbol{c}}\left(\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}\right) = \frac{2\boldsymbol{G}_1(\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c})\boldsymbol{c} - 2\boldsymbol{G}_2(\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c})\boldsymbol{c}}{(\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c})^2} = \boldsymbol{0}$$

Hence,

$$\boldsymbol{G}_1\boldsymbol{c} - \frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}\boldsymbol{G}_2\boldsymbol{c} = \boldsymbol{0}.$$

We then get

$$\left[\boldsymbol{G}_1 - e_{max}\left(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1\right)\boldsymbol{G}_2\right]\boldsymbol{c}^* = \boldsymbol{0},$$

which can be written as

$$\left[\boldsymbol{G}_2^{-1}\boldsymbol{G}_1 - e_{max}\left(\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}\right)\boldsymbol{I}_r\right]\boldsymbol{c}^* = \boldsymbol{0}.$$

Thus, $\boldsymbol{c}^*$ is an eigenvector of $\boldsymbol{G}_2^{-1}\boldsymbol{G}_1$ corresponding to $e_{max}\left(\frac{\boldsymbol{c}'\boldsymbol{G}_1\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_2\boldsymbol{c}}\right)$. The elements of $\boldsymbol{c}^*$ produce the linear combination

$$\boldsymbol{y}_{\boldsymbol{c}^*} = \sum_{i=1}^{r} c_i^*\boldsymbol{y}_i.$$

14.7 In this case, $n = 21$, $\nu_{PE} = 4$, $\nu_{LOF} = n - \rho - \nu_{PE} = 9$. The value of Roy's largest root test statistic is $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1) = 6.3467$. The 10% critical values from using Foster (1957) tables is $\lambda_{0.10} = \frac{x_{0.10}}{1-x_{0.10}} = \frac{0.9922}{1-0.9922} = 127.205$. Therefore, no significant lack of fit can be detected at the 10% level.

14.8 Using Exercise 14.6, model (14.19) is inadequate for a value of $\boldsymbol{c}$ equal to an eigenvector of $\boldsymbol{G}_2^{-1}\boldsymbol{G}_1$ corresponding to $e_{max}(\boldsymbol{G}_2^{-1}\boldsymbol{G}_1) = 6.3467$. This eigenvector is given by $(0.3636, -0.8499, -0.3814)'$. We thus have the linear combination

$$y_c = 0.3636y_1 - 0.8499y_2 - 0.3814y_3.$$

CHAPTER 15

15.1 **(a)** $\boldsymbol{A}$ is nonsingular since its determinant, $|\boldsymbol{A}|$, is equal to 9 which is not equal to zero.

(b) The eigenvalues of $\boldsymbol{A}$ are 6.411, 1.815, 0.773, which are positive. Since $\boldsymbol{A}$ is symmetric, then it must be positive definite by Theorem 6.6.

(c) Since $\boldsymbol{A}$ is symmetric, then its rank is equal to the number of nonzero eigenvalues (see Section 6.5.4), hence $\boldsymbol{A}$ is of rank 3. The matrix $\boldsymbol{B} = \boldsymbol{A}^2$ has three nonzero eigenvalues, namely, 41.107, 3.295, 0.598, hence $\boldsymbol{B}$ is of rank 3

15.2 **(a)** The inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & -0.3333 & -0.3333 \\ -0.3333 & 0.3333 & 0 \\ -0.3333 & 0 & 0.6667 \end{bmatrix}.$$

The trace of $\boldsymbol{A}^2$ is equal to 45.

(b) $e_{max}(\boldsymbol{A}) = 6.41147$, $e_{min}(\boldsymbol{A}) = 0.7733$, $m = 3$, $s = 2.4495$, then $e_{max}(\boldsymbol{A}) \leq m + s\sqrt{2}$ is satisfied since $6.41147 \leq 3 + 2.4495\sqrt{2} = 6.464$.

(c) $e_{max}(\boldsymbol{A}) - e_{min}(\boldsymbol{A}) \leq s\sqrt{6}$ is satisfied since $5.6382 \leq 6.00003$.

15.3 **(a)** See Figure 1.

(b) See Figure 2.

(c) In both plots, the residuals show a random scatter of points contained in a horizontal band, which do not indicate any obvious model defects or inadequacies.

15.4 The points do not deviate significantly from a straight line (see Figure 3). Therefore, there is no indication of a severe problem with the normality assumption.

15.5 **(a)** $tr(\boldsymbol{A}) = 31$ and the eigenvalues of $\boldsymbol{A}$ are 19.961465, 12.634171, 3.3036818, -4.899318. Hence, $\sum_{i=1}^{4} \lambda_i = 31$.

(b) $|\boldsymbol{A}| = -4082$, and $\prod_{i=1}^{4} \lambda_i = -4082$.

15.6 Cook's D_i values for all the observations are less than one, so none of these observations are highly influential. The largest D_i value is 0.247 for observation 9 with a studentized residual of 1.614, so it is mildly influential.

15.7 **(a)** The matrix $\boldsymbol{B}$ is symmetric and its eigenvalues, namely, 0.1628, 5.7527, 17.0845, are positive, hence $\boldsymbol{B}$ is positive definite by Theorem 6.6.

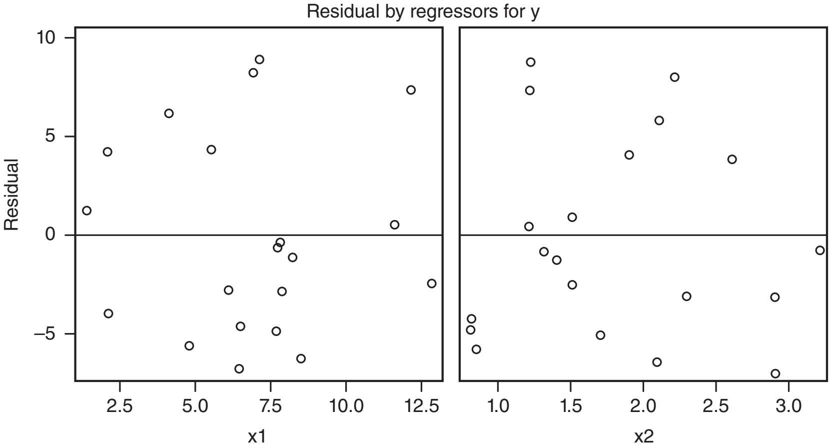


Figure 1 *Plot of the Residuals Against the Regressors.*

(b) By Corollary 9.3,

$$e_{min}(\boldsymbol{B}^{-1}\boldsymbol{A}) \leq \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}} \leq e_{max}(\boldsymbol{B}^{-1}\boldsymbol{A}).$$

The eigenvalues of $\boldsymbol{B}^{-1}\boldsymbol{A}$ are -3.0203, -0.0842, 1.2294. Hence, $e_{min}(\boldsymbol{B}^{-1}\boldsymbol{A}) = -3.0203$, $e_{max}(\boldsymbol{B}^{-1}\boldsymbol{A}) = 1.2294$.

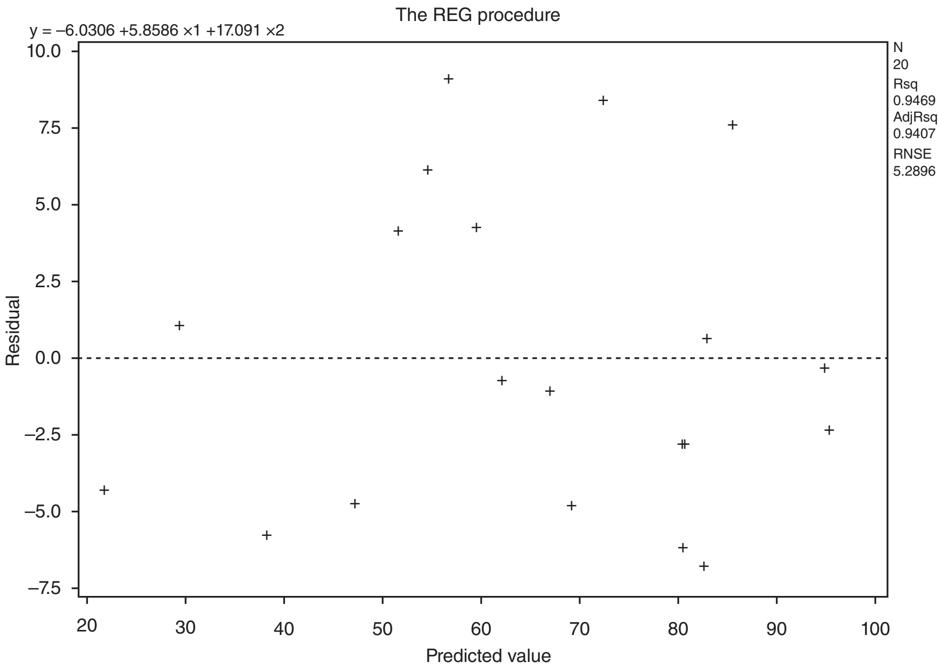


Figure 2 *Plot of the Residuals Against the Predicted Values.*

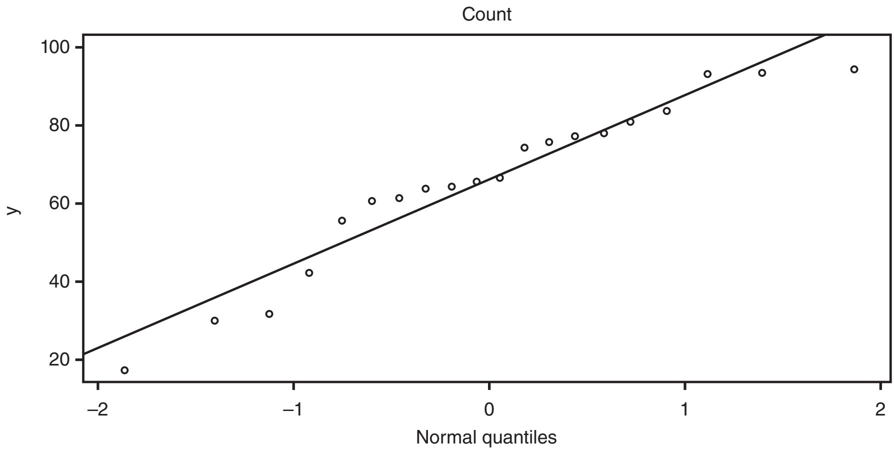


Figure 3 *Plot of Normal Quantiles.*

15.8 **(a)** $e_{min}(\boldsymbol{A}) = -4.1028$, $m - s\sqrt{n-1} = -5.8159$, $m - \frac{s}{\sqrt{n-1}} = -2.2413$, and $-5.8159 \leq -4.1028 \leq -2.2413$.

(b) $e_{max}(\boldsymbol{A}) = 8.0726$. $m + \frac{s}{\sqrt{n-1}} = 4.9079$, $m + s\sqrt{n-1} = 8.4825$, and $4.9079 \leq 8.0726 \leq 8.4825$.

(c) $e_{max}(\boldsymbol{A}) - e_{min}(\boldsymbol{A}) = 12.1754$, $s\sqrt{2n} = 5.0553(\sqrt{6}) = 12.3829$. $e_{max}(\boldsymbol{A}) - e_{min}(\boldsymbol{A}) \leq s\sqrt{2n}$ is true since $12.1754 \leq 12.3829$.

15.9 $[tr(\boldsymbol{A}'\boldsymbol{B})]^2 = 2601$, $tr(\boldsymbol{A}'\boldsymbol{A}) = 215$, $tr(\boldsymbol{B}'\boldsymbol{B}) = 203$, $tr(\boldsymbol{A}'\boldsymbol{A})tr(\boldsymbol{B}'\boldsymbol{B}) = 43645$. The inequality is true since $2601 \leq 215 \times 203$.

15.10 **(a)** If $\boldsymbol{A} = (a_{ij})$, then $tr(\boldsymbol{AA}') = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$, $tr(\boldsymbol{A}^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ji}$. We also have $\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij} - a_{ji})^2 \geq 0$. Then, $\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}^2 - 2a_{ij}a_{ji} + a_{ji}^2) \geq 0$. But, $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ji}^2$. We therefore have $2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ji} \geq 0$, which implies $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 \geq \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ji}$. We thus have $tr(\boldsymbol{AA}') \geq tr(\boldsymbol{A}^2)$.

(b) $tr(\boldsymbol{AA}') = 194$, $tr(\boldsymbol{A}^2) = 118$.

15.11 **(a)** Let $\boldsymbol{C} = \boldsymbol{AB} - \boldsymbol{BA}$. Then, $\boldsymbol{C}' = \boldsymbol{BA} - \boldsymbol{AB} = -\boldsymbol{C}$. We thus have

$$\begin{aligned} tr(\boldsymbol{CC}') &= tr[(\boldsymbol{AB} - \boldsymbol{BA})(\boldsymbol{BA} - \boldsymbol{AB})] \\ &= tr[\boldsymbol{ABBA} - (\boldsymbol{AB})^2 - \boldsymbol{BABA} + \boldsymbol{BA}^2\boldsymbol{B}] \\ &= 2tr(\boldsymbol{A}^2\boldsymbol{B}^2) - 2tr[(\boldsymbol{AB})^2] \geq 0. \end{aligned}$$

Since $tr(\boldsymbol{CC}') = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}^2 \geq 0$, we then conclude that $tr(\boldsymbol{A}^2\boldsymbol{B}^2) \geq tr[(\boldsymbol{AB})^2]$.

(b) $tr[(\boldsymbol{AB})^2] = 13691$, $tr(\boldsymbol{A}^2\boldsymbol{B}^2) = 14928$.

15.12 **(a)** Since $\boldsymbol{A}$ is positive definite, then its eigenvalues, λ_i $(i = 1, 2, \ldots, n)$ are positive. We have $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i$, $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$. We also have that

$$\left(\prod_{i=1}^{n} \lambda_i\right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^{n} \lambda_i.$$

The left hand side is the geometric mean of the λ_i's which is less than or equal to the arithmetic mean, $\frac{1}{n} \sum_{i=1}^{n} \lambda_i$.

(b) $|\boldsymbol{A}|^{1/3} = 1.2599$, $\frac{1}{3} tr(\boldsymbol{A}) = 6$. The eigenvalues of $\boldsymbol{A}$ are $12.1881, 5.7835, 0.0284$ which are positive. Hence, $\boldsymbol{A}$ is positive definite.

CHAPTER 16

16.1 **(a)** $r(\boldsymbol{A}) = 3$, therefore $\boldsymbol{A}$ is nonsingular.

(b) $det(\boldsymbol{A}) = 32.0$.

(c) The eigenvalues are 0.2736, 5.8990, 19.8274.

(d) $tr(\boldsymbol{A}) = 26$.

16.2 **(a)** The eigenvalues of $\boldsymbol{A}$ are 1.1716, 4.0, 6.8284 which are positive. Hence, $\boldsymbol{A}$, which is symmetric, is positive definite by Theorem 6.6.

(b) $|\boldsymbol{A}| = 32$, $\prod_{i=1}^{3} a_{ii} = 64$. Hence, $|\boldsymbol{A}| \leq \prod_{i=1}^{3} a_{ii}$.

16.3 **(a)** The leading principal minors are 4, $\left|\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}\right| = 12$, $|\boldsymbol{A}| = 32$, which are all positive.

(b) The eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ are $1.3726, 16.0, 46.6274$ which are all positive. Hence, $\boldsymbol{A}'\boldsymbol{A}$ is positive definite.

16.4 **(a)** The eigenvalues of $\boldsymbol{A}$ are $0, 14, 53$, which are non-negative. Hence, $\boldsymbol{A}$ is positive semidefinite.

(b) We have by the spectral decomposition theorem, $\boldsymbol{A} = \boldsymbol{PDP}'$, where the diagonal elements of $\boldsymbol{D}$ are the eigenvalues of $\boldsymbol{A}$ and the columns of $\boldsymbol{P}$ are orthonormal eigenvectors of $\boldsymbol{A}$. Hence,

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{P} * [diag(0, 14^{1/2}, 53^{1/2})] * \boldsymbol{P}' \\ &= \begin{bmatrix} 5.1223 & -0.5345 & -3.2367 \\ -0.5345 & 3,4744 & -0.8018 \\ -3.2367 & -0.8018 & 2.4251 \end{bmatrix}.\end{aligned}$$

16.5 **(a)** The inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 0.375 & 0.25 & 0.125 \\ 0.25 & 0.50 & 0.25 \\ 0.125 & 0.25 & 0.375 \end{bmatrix}$$

(b) $tr(\boldsymbol{A})tr(\boldsymbol{A}^{-1}) = 15$. Hence, $tr(\boldsymbol{A})tr(\boldsymbol{A}^{-1}) \geq 9$.

16.6 **(a)** We have that $\boldsymbol{A}' = -\boldsymbol{A}$. Hence, $\boldsymbol{A}$ is skew-symmetric.

(b) We have that

$$(\boldsymbol{I}_3 - \boldsymbol{A})(\boldsymbol{I}_3 + \boldsymbol{A})^{-1} = \begin{bmatrix} 0.5294 & -0.4706 & -0.7059 \\ -0.2824 & -0.8824 & 0.3765 \\ -0.80 & 0 & -0.6 \end{bmatrix}$$

It can be verified that $\boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I}_3$, hence $\boldsymbol{B}$ is orthogonal.

16.7 **(a)** $\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}$, where

$$\boldsymbol{T} = \begin{bmatrix} 1.4142 & 1.4142 & 0.7071 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 1.2247 \end{bmatrix}.$$

(b) $\boldsymbol{A} = \boldsymbol{PDP}'$, where $\boldsymbol{D}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$, namely, 0.7733, 1.8152, 6.4115, and $\boldsymbol{P}$ is the matrix of orthonormal eigenvectors of $\boldsymbol{A}$, namely,

$$\boldsymbol{P} = \begin{bmatrix} 0.4491 & 0.2931 & 0.8440 \\ 0.8440 & -0.4491 & -0.2931 \\ 0.2931 & 0.8440 & -0.4491 \end{bmatrix}.$$

16.8 **(a)** The Moore–Penrose inverse of $\boldsymbol{A}$ is

$$\boldsymbol{B} = \begin{bmatrix} -0.1097 & 0.1248 & 0.0770 \\ 0.1419 & -0.0868 & -0.0299 \end{bmatrix}.$$

(b) The two equalities have been verified.

16.9 **(a)** The Moore–Penrose inverse of $\boldsymbol{A}$ is

$$\boldsymbol{B} = \begin{bmatrix} 0.0687 & 0.0603 & 0.0518 \\ 0.0386 & 0.0019 & -0.0348 \\ 0.1337 & 0.0716 & 0.0094 \\ 0.0508 & -0.0057 & -0.0621 \end{bmatrix}$$

It can be verified that $\boldsymbol{BABA} = \boldsymbol{BA}$.

(b) The eigenvalues of $\boldsymbol{BA}$ are 1, 1, 0, 0.

(c) The trace of $\boldsymbol{BA}$ is 2.

(d) The rank of $\boldsymbol{BA}$ is 2 since $\boldsymbol{BA}$ is idempotent whose trace is 2.

16.10 $[tr(\boldsymbol{A}'\boldsymbol{B})]^2 = 10816$.
$tr(\boldsymbol{A}'\boldsymbol{A}) = 349$.
$tr(\boldsymbol{B}'\boldsymbol{B}) = 339$.
$tr(\boldsymbol{A}'\boldsymbol{A})\, tr(\boldsymbol{B}'\boldsymbol{B}) = 118311$.
Hence, $[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \leq tr(\boldsymbol{A}'\boldsymbol{A})\, tr(\boldsymbol{B}'\boldsymbol{B})$.

16.11 **(a)** The singular-value decomposition of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{PRQ}'$, where

$$\boldsymbol{P} = \begin{bmatrix} -0.6 & 0.8 \\ 0.8 & 0.6 \end{bmatrix},$$

$$\boldsymbol{R} = \begin{bmatrix} 17.3205 & 0 & 0 \\ 0 & 7.0711 & 0 \end{bmatrix},$$

$$\boldsymbol{Q} = \begin{bmatrix} -0.5774 & 0.7071 & -0.4082 \\ -0.5774 & -0.7071 & -0.4082 \\ -0.5774 & 0 & 0.8165 \end{bmatrix}.$$

(b) The singular values of $\boldsymbol{A}$ are 17.3205 and 7.70711. The positive eigenvalues of $\boldsymbol{AA}'$ are 50 and 300. Their positive square roots are equal to the singular values of $\boldsymbol{A}$.

16.12 **(a)** The direct product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} 4 & 0 & -2 & 0 & 0 & 0 \\ 12 & 16 & -6 & -8 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 & 0 \\ -6 & -8 & 12 & 16 & -6 & -8 \\ 0 & 0 & -2 & 0 & 4 & 0 \\ 0 & 0 & -6 & -8 & 12 & 16 \end{bmatrix}.$$

(b) $(\boldsymbol{A} \otimes B)^{-1} - \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1} = \boldsymbol{0}$ was verified.

16.13 **(a)** $tr[(\boldsymbol{AB})^2] = 1951$, $tr(\boldsymbol{A}^2\boldsymbol{B}^2) = 7764$. This verifies the inequality in (a).

(b) $|\boldsymbol{A}'\boldsymbol{B}|^2 = 189,888,400$, $|\boldsymbol{A}'\boldsymbol{A}| = 70225$, $|\boldsymbol{B}'\boldsymbol{B}| = 2704$. Hence, $|\boldsymbol{A}'\boldsymbol{A}|\ |\boldsymbol{B}'\boldsymbol{B}| = 189,888,400$. This verifies the inequality in (b).

16.14 The MATLAB commands for this plot are
$x = 0 : pi/100 : 2 * pi;$
$y = (sin(x)).{}^*cos(x) + cos(3 * x);$
$plot(x, y)$

The plot is shown in Figure 4.

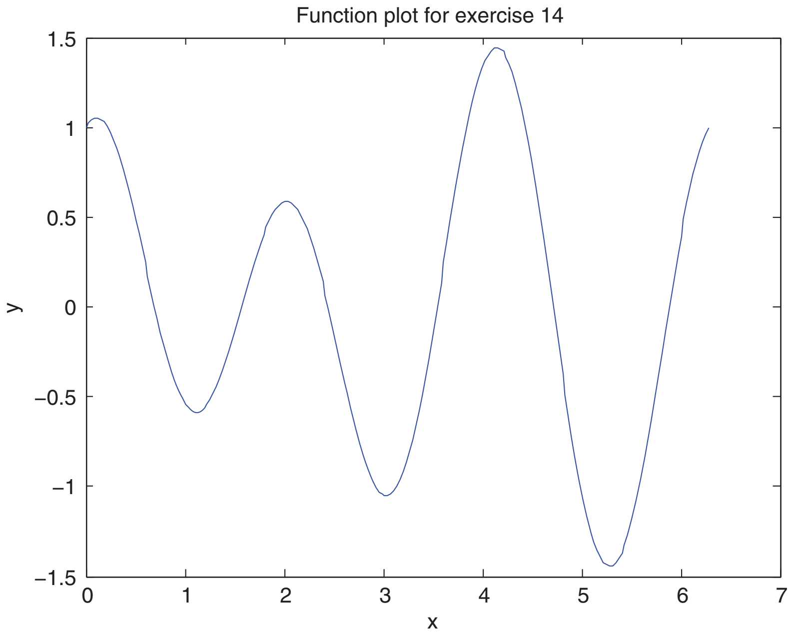


Figure 4 *Plot of the y Function.*

16.15 The MATLAB commands for this plot are
$[x, y] = meshgrid(-2 : .2 : 2, -2 : .2 : 2);$
$z = y. * exp(-3. * x.^2 - 7. * y.^2);$
$surf(z)$

The plot is shown in Figure 5.

CHAPTER 17

17.1 **(a)** The rank is obtained by using the function $qr(A)\$rank$ which gives the value 3.
(b) $det(\boldsymbol{A}) = 130$.
(c) $eigen(\boldsymbol{A})$ gives the following eigenvalues of $\boldsymbol{A}$: 18.359257, 5.306314, 1.334428.
(d) The trace is given by $sum(diag(\boldsymbol{A})) = 25$.

17.2 **(a)** $eigen(\boldsymbol{A})$ gives the following eigenvalues of $\boldsymbol{A}$: 6.414214, 5.0, 3.585786. Since they are positive and $\boldsymbol{A}$ is symmetric, it must be positive definite.
(b) The inverse of $\boldsymbol{A}$ is given by the function $solve(A)$ which gives the matrix,

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 0.208696 & 0.043478 & 0.008696 \\ 0.043478 & 0.217391 & 0.043478 \\ 0.008696 & 0.043478 & 0.208696 \end{bmatrix}.$$

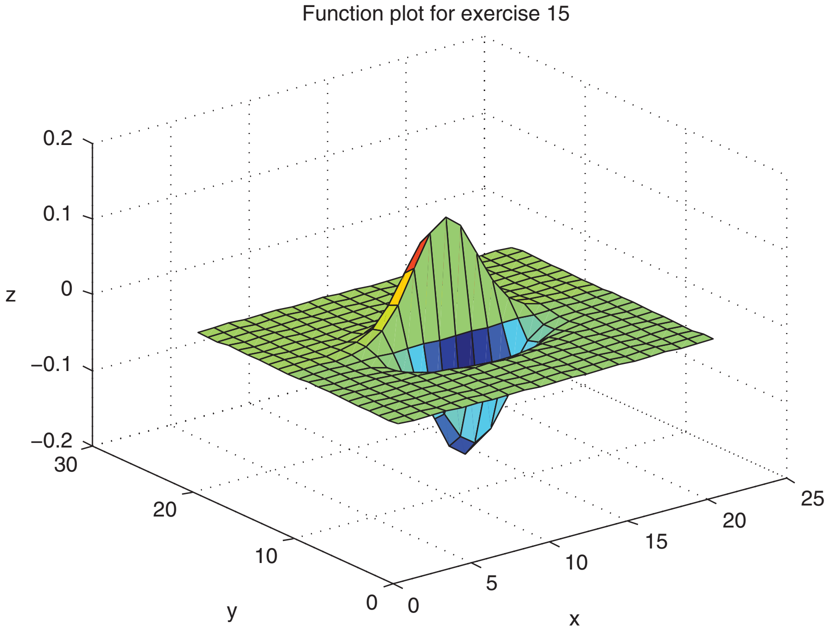


Figure 5 *Plot of the z Function.*

(c) The trace is given by $sum(diag(\boldsymbol{A})) = 15$, which is also equal to the sum of eigenvalues of $\boldsymbol{A}$.

17.3 **(a)** The leading principal minors are 5, the 2×2 determinant of $\boldsymbol{B}$ given by

$$\boldsymbol{B} = \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix},$$

which has the value 24, and the determinant of $\boldsymbol{A}$ which has the value, $det(A) = 115$.

(b) We use the product $t(\boldsymbol{A})\% * \%\boldsymbol{A}$ to get the matrix $\boldsymbol{A}'\boldsymbol{A}$. Its eigenvalues are given by 41.14214, 25, 12.85786, which are positive. Hence, $\boldsymbol{A}$ is positive definite.

17.4 **(a)** Using $eigen(\boldsymbol{A})$, the eigenvalues of $\boldsymbol{A}$ are 0, 14, 53, which are non-negative. Hence, $\boldsymbol{A}$ is positive semidefinite by Theorem 6.7.

(b) The leading principal minors are 37, the determinant of the leading principal submatrix of order 2, denoted by $\boldsymbol{B}$,

$$\boldsymbol{B} = \begin{bmatrix} 37 & -2 \\ -2 & 13 \end{bmatrix},$$

which is equal to 477, and the determinant of $\boldsymbol{A}$, $det(\boldsymbol{A}) = 0$.

(c) Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ denote two principal submatrices of $\boldsymbol{A}$ of order 2 given by

$$\boldsymbol{A}_1 = \begin{bmatrix} 37 & -24 \\ -24 & 17 \end{bmatrix},$$

$$\boldsymbol{A}_2 = \begin{bmatrix} 13 & -3 \\ -3 & 17 \end{bmatrix}.$$

The values of two principal minors of order 2 are $|\boldsymbol{A}_1| = 53$ and $|\boldsymbol{A}_2| = 212$.

17.5 **(a)** Let $\boldsymbol{B} = \boldsymbol{A}^2$, that is, $\boldsymbol{B} = \boldsymbol{A}\% * \%\boldsymbol{A}$, then the trace of $\boldsymbol{A}^2$ is $sum(diag(\boldsymbol{B})) = 3005$.

(b) $tr(\boldsymbol{A}) = sum(diag(\boldsymbol{A})) = 67$, and its square is 4489. Hence, $tr(\boldsymbol{A}^2) \leq [tr(\boldsymbol{A})]^2$.

17.6 **(a)** It is skew-symmetric since $t(\boldsymbol{A}) = -\boldsymbol{A}$.

(b) Let $\boldsymbol{E}1 = diag(3) - \boldsymbol{A}, \boldsymbol{E}2 = diag(3) + \boldsymbol{A}$, and $\boldsymbol{E} = \boldsymbol{E}1\% * \%solve(\boldsymbol{E}2)$. Hence,

$$\boldsymbol{E} = \begin{bmatrix} 0.529412 & -0.470588 & -0.705882 \\ -0.282353 & -0.882353 & 0.376471 \\ -0.80 & 0 & -0.60 \end{bmatrix}.$$

It follows that $\boldsymbol{E}\% * \%t(\boldsymbol{E}) = \boldsymbol{I}_3$, that is, $\boldsymbol{E}$ is orthogonal.

17.7 **(a)** The Cholesky decomposition of $\boldsymbol{A}$ is obtained in R using the function $chol(\boldsymbol{A})$ which gives the upper triangular matrix $\boldsymbol{T}$ so that $\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}$, where $\boldsymbol{T}$ is given

$$\boldsymbol{T} = \begin{bmatrix} 1.414214 & 1.414214 & 0.707107 \\ 0.0 & 1.732051 & 0.0 \\ 0.0 & 0.0 & 1.224745 \end{bmatrix}.$$

(b) Using the command $eigen(\boldsymbol{A})$ gives the eigenvalues and corresponding orthonormal eigenvectors of $\boldsymbol{A}$. This provides the spectral decomposition of $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{P\Lambda P}'$, as shown in Theorem 6.2, where

$$\boldsymbol{\Lambda} = \begin{bmatrix} 6.411474 & 0.0 & 0.0 \\ 0.0 & 1.815208 & 0.0 \\ 0.0 & 0.0 & 0.773318 \end{bmatrix},$$

$$\boldsymbol{P} = \begin{bmatrix} -0.449099 & 0.293128 & 0.844030 \\ -0.844030 & -0.449099 & -0.293128 \\ -0.293128 & 0.844030 & -0.449099 \end{bmatrix}.$$

17.8 **(a)** Using $eigen(\boldsymbol{C})$, where $\boldsymbol{C} = t(\boldsymbol{A})\% * \%\boldsymbol{A}$, the eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ are 45, 322.

(b) The determinant $det(\boldsymbol{C})$ is equal to 14490. The diagonal elements of $\boldsymbol{C}$ are 126, 241. Their product is 30366. Hence, the stated inequality is verified.

(c) Using the command $svd(\boldsymbol{A})$, the singular values of $\boldsymbol{A}$ are 6.708204, 17.944358, which give the diagonal elements of $\boldsymbol{D}_1$ in the expression (17.2). In addition, we get the matrices $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$, denoted in R as \$u and \$v, respectively, which are given by

$$\boldsymbol{P}_1 = \begin{bmatrix} -0.458725 & 0.806114 \\ -0.652930 & -0.591150 \\ -0.602705 & 0.026870 \end{bmatrix},$$

$$\boldsymbol{Q}_1 = \begin{bmatrix} -0.540758 & 0.841179 \\ 0.841179 & 0.540758 \end{bmatrix}.$$

To check the expression in (17.2), use the product $svd(\boldsymbol{A})\$u\% * \%diag(svd(\boldsymbol{A})\$d)\% * \%t(svd(\boldsymbol{A})\$v)$, which should give the matrix $\boldsymbol{A}$.

17.9 **(a)** $qr(\boldsymbol{A})\$rank$ gives the rank, namely, 3.

(b) Using the statements, $qra = qr(\boldsymbol{A})$, $\boldsymbol{Q} = qr.Q(qra)$, $\boldsymbol{R} = qr.R(qra)$, we get

$$\boldsymbol{Q} = \begin{bmatrix} -0.169031 & 0.897085 & 0.408248 \\ -0.507093 & 0.276026 & -0.816497 \\ -0.845154 & -0.345033 & 0.408248 \end{bmatrix},$$

$$\boldsymbol{R} = \begin{bmatrix} -5.91608 & 3.549648 & -6.592203 \\ 0.0 & 2.898275 & 8.41880 \\ 0.0 & 0.0 & -4.082483 \end{bmatrix}.$$

(c) $\boldsymbol{Q}\% * \%\boldsymbol{R} = \boldsymbol{A}$.

17.10 Let $\boldsymbol{C} = t(\boldsymbol{A})\% * \%\boldsymbol{B}$. Its trace is $t = sum(diag(\boldsymbol{C})) = 104$. Also, let $\boldsymbol{D} = t(\boldsymbol{A})\% * \%\boldsymbol{A}$, $\boldsymbol{E} = t(\boldsymbol{B})\% * \%\boldsymbol{B}$. Their traces are $u = sum(diag(\boldsymbol{D})) = 349$, $v = sum(diag(\boldsymbol{E})) = 339$. Hence, $t^2 = 10816 \leq u * v = 118311$.

17.11 **(a)** $||\boldsymbol{A}||_2^2 = tr(\boldsymbol{A}^2) = sum(diag(\boldsymbol{B})) = 573$, where $\boldsymbol{B} = \boldsymbol{A}^2$.

(b) $eigen(\boldsymbol{A})$ gives the eigenvalues -14.100689, 7.112134, 17.988555. Their sum of squares is 573.

17.12 **(a)** Let $\boldsymbol{C}$ denote the direct product $\boldsymbol{C} = \boldsymbol{A}\% \times \%\boldsymbol{B}$. Then,

$$\boldsymbol{C} = \begin{bmatrix} 4 & 0 & -2 & 0 & 0 & 0 \\ 12 & 16 & -6 & -8 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 & 0 \\ -6 & -8 & 12 & 16 & -6 & -3 \\ 0 & 0 & -2 & 0 & 4 & 0 \\ 0 & 0 & -6 & -8 & 12 & 16 \end{bmatrix}.$$

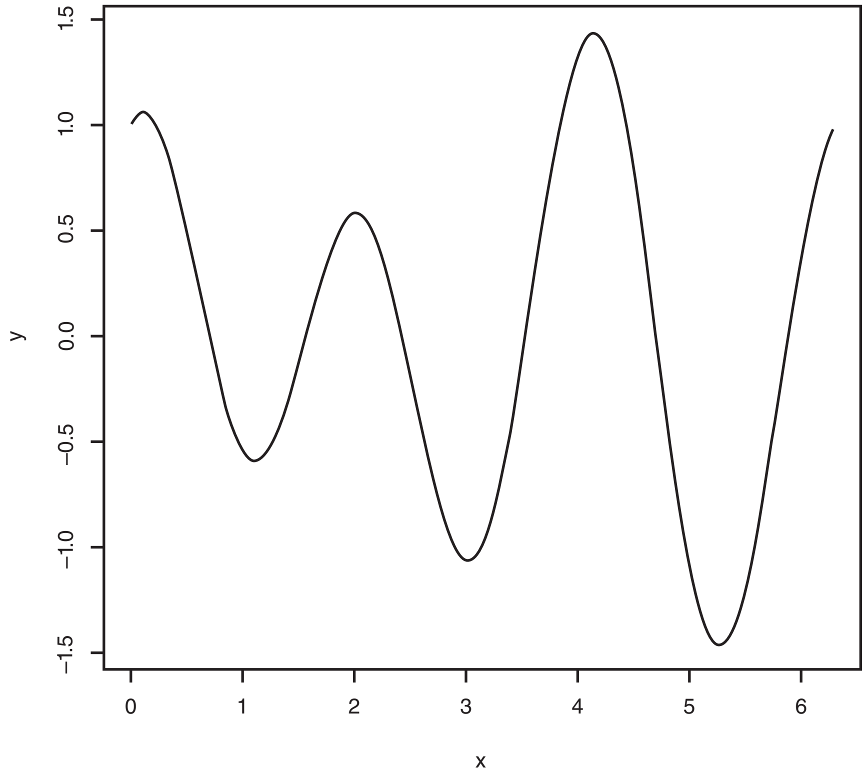


Figure 6 *Plot of the Trigonometric Function.*

(b) $|\boldsymbol{C}| = 65536$, $|\boldsymbol{A}| = 32$, $|\boldsymbol{B}| = 4$. Then, $|\boldsymbol{A}|^2|\boldsymbol{B}|^3 = 65536$. This verifies the given equality.

17.13 **(a)** Let $\boldsymbol{C} = \boldsymbol{A}\% * \%\boldsymbol{B}, \boldsymbol{D} = \boldsymbol{C}\% * \%\boldsymbol{C}$. Then, $t1 = sum(diag(\boldsymbol{D})) = 1951$. Let $\boldsymbol{E} = \boldsymbol{A}\% * \%\boldsymbol{A}\% * \%\boldsymbol{B}\% * \%\boldsymbol{B}$, $t2 = sum(diag(\boldsymbol{E})) = 7764$. Hence, $t1 \leq t2$.

(b) Let $t3 = sum(diag(\boldsymbol{A}\% * \%\boldsymbol{B})) = 29$, $t4 = sum(diag(\boldsymbol{A}\% * \%\boldsymbol{A})) = 224$, $t5 = sum(diag(\boldsymbol{B}\% * \%\boldsymbol{B})) = 93$. We conclude that $t3 \leq \frac{1}{2}(t4 + t5) = 158.5$.

17.14 The needed R statements for the plotting of this function are
x=seq(0,6.28,by=0.01)
y=sin(x)*cos(x)+cos(3*x)
plot(x,y,type="l",xlab="x",ylab="y")

The plot is shown in Figure 6.

17.15 The R statements for this three-dimensional plot are
x=seq(-2,2,0.1)
y=seq(-2,2,0.1)
$z = outer(x, y, FUN = function(x, y)(y/(1 + 3 * x^2 + 7 * y^2)))$
persp(x,y,z,theta=40,phi=40,ticktype="detailed")

The plot is shown in Figure 7.

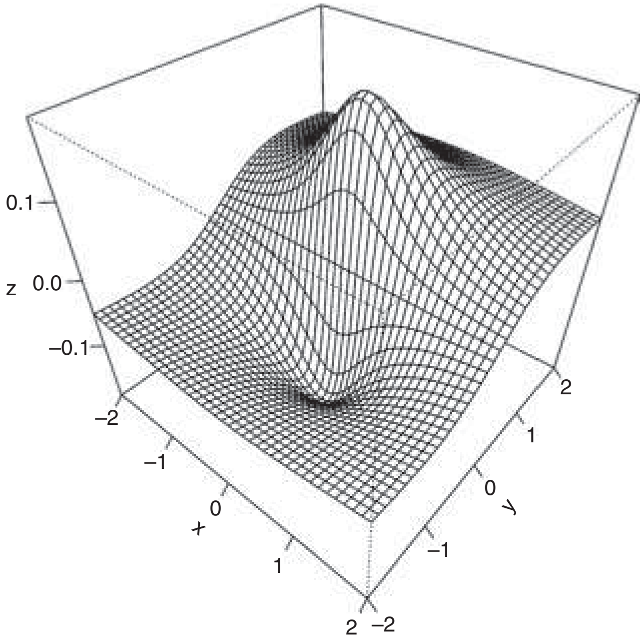


Figure 7 *Plot of a Three-Dimensional Figure.*

Index

Matrix Algebra Useful for Statistics, Second Edition. Shayle R. Searle and André I. Khuri.

Companion website: www.wiley.com/go/searle/matrixalgebra2e