Wiley Series in Probability and Statistics

Shayle R. Searle and André I. Khuri

Matrix Algebra Useful for Statistics

Second Edition

WILEY

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.

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

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

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 xvii

PREFACE TO THE FIRST EDITION xix

INTRODUCTION xxi

ABOUT THE COMPANION WEBSITE xxxi

PART I DEFINITIONS, BASIC CONCEPTS, AND MATRIX OPERATIONS 1

1 Vector Spaces, Subspaces, and Linear Transformations 3

- 1.1 Vector Spaces / 3
  - 1.1.1 Euclidean Space / 3
- 1.2 Base of a Vector Space / 5
- 1.3 Linear Transformations / 7
  - 1.3.1 The Range and Null Spaces of a Linear Transformation / 8
- Reference / 9
- Exercises / 9

2 Matrix Notation and Terminology 11

- 2.1 Plotting of a Matrix / 14
- 2.2 Vectors and Scalars / 16
- 2.3 General Notation / 16
- Exercises / 17

3 Determinants **21**

3.1 Expansion by Minors / 21
3.1.1 First- and Second-Order Determinants / 22
3.1.2 Third-Order Determinants / 23
3.1.3 n-Order Determinants / 24
3.2 Formal Definition / 25
3.3 Basic Properties / 27
3.3.1 Determinant of a Transpose / 27
3.3.2 Two Rows the Same / 28
3.3.3 Cofactors / 28
3.3.4 Adding Multiples of a Row (Column) to a Row (Column) / 30
3.3.5 Products / 30
3.4 Elementary Row Operations / 34
3.4.1 Factorization / 35
3.4.2 A Row (Column) of Zeros / 36
3.4.3 Interchanging Rows (Columns) / 36
3.4.4 Adding a Row to a Multiple of a Row / 36
3.5 Examples / 37
3.6 Diagonal Expansion / 39
3.7 The Laplace Expansion / 42
3.8 Sums and Differences of Determinants / 44
3.9 A Graphical Representation of a 3×3 Determinant / 45
References / 46
Exercises / 47

4 Matrix Operations **51**

4.1 The Transpose of a Matrix / 51
4.1.1 A Reflexive Operation / 52
4.1.2 Vectors / 52
4.2 Partitioned Matrices / 52
4.2.1 Example / 52
4.2.2 General Specification / 54
4.2.3 Transposing a Partitioned Matrix / 55
4.2.4 Partitioning Into Vectors / 55
4.3 The Trace of a Matrix / 55
4.4 Addition / 56
4.5 Scalar Multiplication / 58
4.6 Equality and the Null Matrix / 58
4.7 Multiplication / 59
4.7.1 The Inner Product of Two Vectors / 59
4.7.2 A Matrix–Vector Product / 60

4.7.3 A Product of Two Matrices / 62
4.7.4 Existence of Matrix Products / 65
4.7.5 Products With Vectors / 65
4.7.6 Products With Scalars / 68
4.7.7 Products With Null Matrices / 68
4.7.8 Products With Diagonal Matrices / 68
4.7.9 Identity Matrices / 69
4.7.10 The Transpose of a Product / 69
4.7.11 The Trace of a Product / 70
4.7.12 Powers of a Matrix / 71
4.7.13 Partitioned Matrices / 72
4.7.14 Hadamard Products / 74
4.8 The Laws of Algebra / 74
4.8.1 Associative Laws / 74
4.8.2 The Distributive Law / 75
4.8.3 Commutative Laws / 75
4.9 Contrasts With Scalar Algebra / 76
4.10 Direct Sum of Matrices / 77
4.11 Direct Product of Matrices / 78
4.12 The Inverse of a Matrix / 80
4.13 Rank of a Matrix—Some Preliminary Results / 82
4.14 The Number of LIN Rows and Columns in a Matrix / 84
4.15 Determination of the Rank of a Matrix / 85
4.16 Rank and Inverse Matrices / 87
4.17 Permutation Matrices / 87
4.18 Full-Rank Factorization / 89
4.18.1 Basic Development / 89
4.18.2 The General Case / 91
4.18.3 Matrices of Full Row (Column) Rank / 91
References / 92
Exercises / 92

5 Special Matrices **97**

5.1 Symmetric Matrices / 97
5.1.1 Products of Symmetric Matrices / 97
5.1.2 Properties of $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ / 98
5.1.3 Products of Vectors / 99
5.1.4 Sums of Outer Products / 100
5.1.5 Elementary Vectors / 101
5.1.6 Skew-Symmetric Matrices / 101
5.2 Matrices Having All Elements Equal / 102

5.3 Idempotent Matrices / 104
5.4 Orthogonal Matrices / 106
5.4.1 Special Cases / 107
5.5 Parameterization of Orthogonal Matrices / 109
5.6 Quadratic Forms / 110
5.7 Positive Definite Matrices / 113
References / 114
Exercises / 114

6 Eigenvalues and Eigenvectors **119**

6.1 Derivation of Eigenvalues / 119
6.1.1 Plotting Eigenvalues / 121
6.2 Elementary Properties of Eigenvalues / 122
6.2.1 Eigenvalues of Powers of a Matrix / 122
6.2.2 Eigenvalues of a Scalar-by-Matrix Product / 123
6.2.3 Eigenvalues of Polynomials / 123
6.2.4 The Sum and Product of Eigenvalues / 124
6.3 Calculating Eigenvectors / 125
6.3.1 Simple Roots / 125
6.3.2 Multiple Roots / 126
6.4 The Similar Canonical Form / 128
6.4.1 Derivation / 128
6.4.2 Uses / 130
6.5 Symmetric Matrices / 131
6.5.1 Eigenvalues All Real / 132
6.5.2 Symmetric Matrices Are Diagonable / 132
6.5.3 Eigenvectors Are Orthogonal / 132
6.5.4 Rank Equals Number of Nonzero Eigenvalues for a Symmetric Matrix / 135
6.6 Eigenvalues of Orthogonal and Idempotent Matrices / 135
6.6.1 Eigenvalues of Symmetric Positive Definite and Positive Semidefinite Matrices / 136
6.7 Eigenvalues of Direct Products and Direct Sums of Matrices / 138
6.8 Nonzero Eigenvalues of *AB* and *BA* / 140
References / 141
Exercises / 141

7 Diagonalization of Matrices **145**

7.1 Proving the Diagonability Theorem / 145
7.1.1 The Number of Nonzero Eigenvalues Never Exceeds Rank / 145

7.1.2 A Lower Bound on $r\left(\mathbf{A} - \lambda_k \mathbf{I}\right)$ / 146
7.1.3 Proof of the Diagonability Theorem / 147
7.1.4 All Symmetric Matrices Are Diagonable / 147
7.2 Other Results for Symmetric Matrices / 148
7.2.1 Non-Negative Definite (n.n.d.) / 148
7.2.2 Simultaneous Diagonalization of Two Symmetric Matrices / 149
7.3 The Cayley–Hamilton Theorem / 152
7.4 The Singular-Value Decomposition / 153
References / 157
Exercises / 157

8 Generalized Inverses **159**

8.1 The Moore–Penrose Inverse / 159
8.2 Generalized Inverses / 160
8.2.1 Derivation Using the Singular-Value Decomposition / 161
8.2.2 Derivation Based on Knowing the Rank / 162
8.3 Other Names and Symbols / 164
8.4 Symmetric Matrices / 165
8.4.1 A General Algorithm / 166
8.4.2 The Matrix $\mathbf{X}'\mathbf{X}$ / 166
References / 167
Exercises / 167

9 Matrix Calculus **171**

9.1 Matrix Functions / 171
9.1.1 Function of Matrices / 171
9.1.2 Matrices of Functions / 174
9.2 Iterative Solution of Nonlinear Equations / 174
9.3 Vectors of Differential Operators / 175
9.3.1 Scalars / 175
9.3.2 Vectors / 176
9.3.3 Quadratic Forms / 177
9.4 Vec and Vech Operators / 179
9.4.1 Definitions / 179
9.4.2 Properties of Vec / 180
9.4.3 Vec-Permutation Matrices / 180
9.4.4 Relationships Between Vec and Vech / 181
9.5 Other Calculus Results / 181
9.5.1 Differentiating Inverses / 181
9.5.2 Differentiating Traces / 182
9.5.3 Derivative of a Matrix with Respect to Another Matrix / 182

9.5.4 Differentiating Determinants / 183
9.5.5 Jacobians / 185
9.5.6 Aitken's Integral / 187
9.5.7 Hessians / 188
9.6 Matrices with Elements That Are Complex Numbers / 188
9.7 Matrix Inequalities / 189
References / 193
Exercises / 194

PART II APPLICATIONS OF MATRICES IN STATISTICS 199

10 Multivariate Distributions and Quadratic Forms 201

10.1 Variance-Covariance Matrices / 202
10.2 Correlation Matrices / 203
10.3 Matrices of Sums of Squares and Cross-Products / 204
10.3.1 Data Matrices / 204
10.3.2 Uncorrected Sums of Squares and Products / 204
10.3.3 Means, and the Centering Matrix / 205
10.3.4 Corrected Sums of Squares and Products / 205
10.4 The Multivariate Normal Distribution / 207
10.5 Quadratic Forms and χ^2-Distributions / 208
10.5.1 Distribution of Quadratic Forms / 209
10.5.2 Independence of Quadratic Forms / 210
10.5.3 Independence and Chi-Squaredness of Several Quadratic Forms / 211
10.5.4 The Moment and Cumulant Generating Functions for a Quadratic Form / 211
10.6 Computing the Cumulative Distribution Function of a Quadratic Form / 213
10.6.1 Ratios of Quadratic Forms / 214
References / 215
Exercises / 215

11 Matrix Algebra of Full-Rank Linear Models 219

11.1 Estimation of β by the Method of Least Squares / 220
11.1.1 Estimating the Mean Response and the Prediction Equation / 223
11.1.2 Partitioning of Total Variation Corrected for the Mean / 225
11.2 Statistical Properties of the Least-Squares Estimator / 226
11.2.1 Unbiasedness and Variances / 226
11.2.2 Estimating the Error Variance / 227
11.3 Multiple Correlation Coefficient / 229

11.4 Statistical Properties under the Normality Assumption / 231
11.5 Analysis of Variance / 233
11.6 The Gauss–Markov Theorem / 234
11.6.1 Generalized Least-Squares Estimation / 237
11.7 Testing Linear Hypotheses / 237
11.7.1 The Use of the Likelihood Ratio Principle in Hypothesis Testing / 239
11.7.2 Confidence Regions and Confidence Intervals / 241
11.8 Fitting Subsets of the x-Variables / 246
11.9 The Use of the $R(.|.)$ Notation in Hypothesis Testing / 247
References / 249
Exercises / 249

12 Less-Than-Full-Rank Linear Models **253**

12.1 General Description / 253
12.2 The Normal Equations / 256
12.2.1 A General Form / 256
12.2.2 Many Solutions / 257
12.3 Solving the Normal Equations / 257
12.3.1 Generalized Inverses of $X'X$ / 258
12.3.2 Solutions / 258
12.4 Expected Values and Variances / 259
12.5 Predicted y-Values / 260
12.6 Estimating the Error Variance / 261
12.6.1 Error Sum of Squares / 261
12.6.2 Expected Value / 262
12.6.3 Estimation / 262
12.7 Partitioning the Total Sum of Squares / 262
12.8 Analysis of Variance / 263
12.9 The $R(\cdot|\cdot)$ Notation / 265
12.10 Estimable Linear Functions / 266
12.10.1 Properties of Estimable Functions / 267
12.10.2 Testable Hypotheses / 268
12.10.3 Development of a Test Statistic for H_0 / 269
12.11 Confidence Intervals / 272
12.12 Some Particular Models / 272
12.12.1 The One-Way Classification / 272
12.12.2 Two-Way Classification, No Interactions, Balanced Data / 273
12.12.3 Two-Way Classification, No Interactions, Unbalanced Data / 276
12.13 The $R(\cdot|\cdot)$ Notation (Continued) / 277

12.14 Reparameterization to a Full-Rank Model / 281
References / 282
Exercises / 282

13 Analysis of Balanced Linear Models Using Direct Products of Matrices 287

13.1 General Notation for Balanced Linear Models / 289
13.2 Properties Associated with Balanced Linear Models / 293
13.3 Analysis of Balanced Linear Models / 298
13.3.1 Distributional Properties of Sums of Squares / 298
13.3.2 Estimates of Estimable Linear Functions of the Fixed Effects / 301
References / 307
Exercises / 308

14 Multiresponse Models 313

14.1 Multiresponse Estimation of Parameters / 314
14.2 Linear Multiresponse Models / 316
14.3 Lack of Fit of a Linear Multiresponse Model / 318
14.3.1 The Multivariate Lack of Fit Test / 318
References / 323
Exercises / 324

PART III MATRIX COMPUTATIONS AND RELATED SOFTWARE 327

15 SAS/IML 329

15.1 Getting Started / 329
15.2 Defining a Matrix / 329
15.3 Creating a Matrix / 330
15.4 Matrix Operations / 331
15.5 Explanations of SAS Statements Used Earlier in the Text / 354
References / 357
Exercises / 358

16 Use of MATLAB in Matrix Computations 363

16.1 Arithmetic Operators / 363
16.2 Mathematical Functions / 364
16.3 Construction of Matrices / 365
16.3.1 Submatrices / 365
16.4 Two- and Three-Dimensional Plots / 371
16.4.1 Three-Dimensional Plots / 374
References / 378
Exercises / 379

17 Use of R in Matrix Computations **383**

17.1 Two- and Three-Dimensional Plots / 396

17.1.1 Two-Dimensional Plots / 397

17.1.2 Three-Dimensional Plots / 404

References / 408

Exercises / 408

APPENDIX **413**

INDEX **475**

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

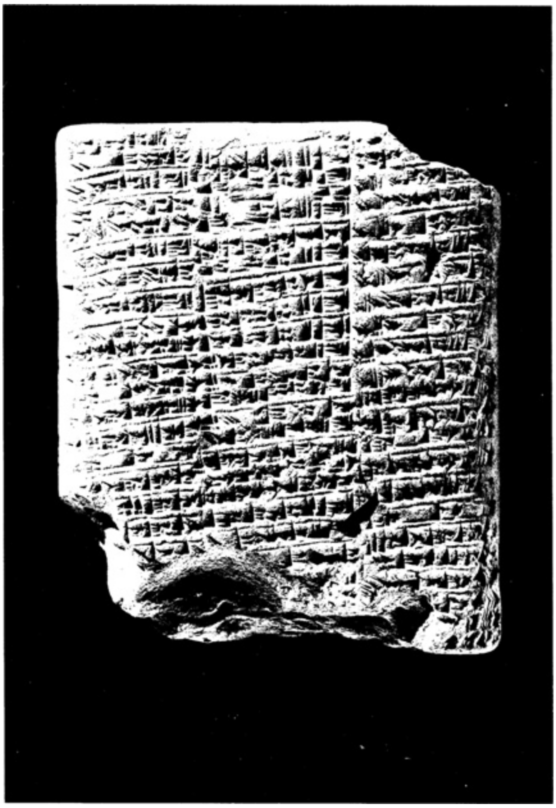

Figure 0.1 *A Cuneiform Tablet with 97 Linear Equations (YBC4695-1). Yale Babylonian Collection, Yale University Library, New Haven, CT.*

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

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

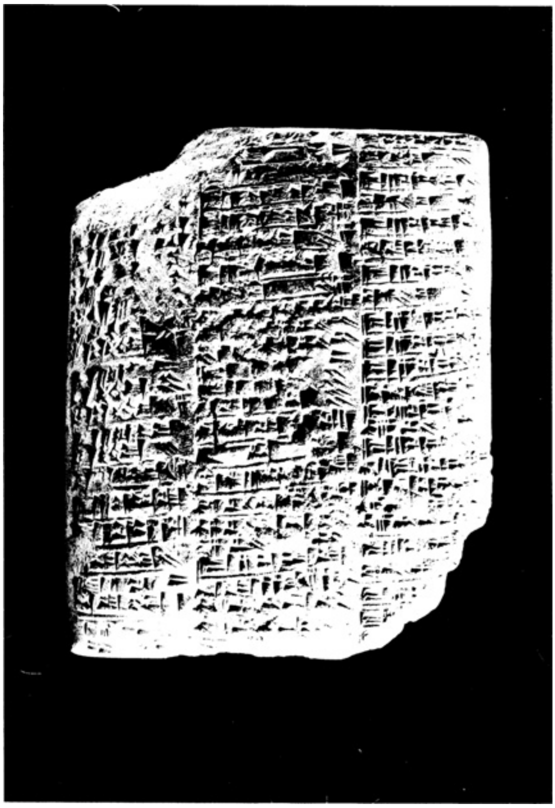

Figure 0.2 *An Old Babylonian Mathematical Text with Linear Equations (YBC4695-2). Yale Babylonian Collection, Yale University Library, New Haven, CT.*

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 $a\,x$, 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.