THE INFUSION OF MATRICES INTO STATISTICS

Shayle R. Searle

Department of Biometrics and Statistical Science
Cornell University
Ithaca, N. Y., 14850, U.S.A.

BU-1444-M March 1999

KEY WORDS: determinants, direct product, Doolittle, experiment design, generalized inverse, multivariate statistics, normal equations, quadratic forms, vec and vech.

1 INTRODUCTION

Matrices were introduced to me as a one-year Master's level 1949 course in New Zealand taught by an A.C. Aitken doctoral graduate. Two years later at Cambridge University's Statistics Laboratory, the 24-week mathematical statistics course made no use of matrices, not even in the teaching of the bivariate normal distribution or of multiple linear regression. This was a great surprise. But it is totally in line with similar comments from others, reported by Farebrother (1997), who also quotes Grattan-Guiness and Lederman (1994) as saying "the rise of matrix theory to staple diet has occurred only since the 1950s".

Indeed, not even Aitken himself (much of whose research centered on both statistics and matrices) made a strong pitch for using matrices in statistics. Neither of his two books (Aitken 1939a and b), *Determinants and Matrices* and *Statistical Mathematics*, mentions the topic of the other, except for a snippet about quadratic forms in the matrix book. For someone having strong interests in both topics these are, surely, unexpected omissions. They have been motivation for my trying to trace a little of the infusion of matrices into statistics. It all seems remarkably recent when viewed against the longer history of matrices themselves. Of course, in trying to be an amateur historian one immediately comes face to face with the ocean of literature available, and the consequent near-impossibility of

assembling every detail and aspect of one's topic. Therefore there are assuredly gaping holes in what follows—and all that can be done is to apologize and ask for help for filling those holes. Circumscribed by such lacework, the paper is arranged under four main headings: origins, early uses, special topics, and books.

2 THE ORIGINS OF MATRICES

In his two-volume *Men of Mathematics*, Bell (1937, 1953, Vol. 2, pp. 441-443), attributes the invention of matrices and their algebra to Cayley (1855, 1858). "Matrices" writes Bell, "grew out of the ... way in which the transformations of the theory of algebraic invariants are combined." For example, on substituting

$$x = \frac{pz+q}{rz+s} \qquad \text{into} \qquad y = \frac{ax+b}{cx+d}$$

each coefficient of z and each non-z term in the resulting expression for y is a term in the matrix product

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix}.$$

However, Farebrother (1997) indicates that others in the eighteenth and nineteenth centuries may well have made greater although indirect contributions than did Cayley, a conclusion, he suggests that is supported by Grattan-Guiness (1994, p. 67).

From that paper, Farebrother (1999) gives the following quote as evidence of the slow adaptation of matrices to statistics.

> "Matrix theory was not to emerge until quite late in the 19'th century and became prominent only from the 1920s. The papers by Farkas (1902), de la Valée Poussin (1911) and Haar (1924) are typical examples of the continuing slowness, for they explicitly worked with determinants but wrote out the matrix equations and inequalities longhand."

Nowadays interest centers much more on matrices than determinants, but this was not always the case. The 4-volume history of determinants by Muir (1890, 1911, 1920 and 1923) attests to this, as does also the 3-volume work of Cullis (1913-1925) who worked with what

he called "determinoids", the extension of determinants to rectangular matrices using the Laplace expansion.

This interest in determinants includes the contention that what is nowadays often called the direct, or Kronecker, product should be called the Zehfuss product, because for $\mathbf{A} \otimes \mathbf{B} = \{a_{ij}\mathbf{B}\}$ Zehfuss (1858) has

$$\text{determinant}(\mathbf{A}_{a\times a} \otimes \mathbf{B}_{b\times b}) = [\text{determinant}(\mathbf{A})]^b[\text{determinant}(\mathbf{B})]^a.$$

Hensel (1891) promotes the name of Kronecker, resulting, no doubt, from his being a student in Berlin when Kronecker lectured there. But Muir (*loc cit.*) espouses the Zehfuss name as do Henderson *et al.* (1983) in their review of this matter. And, of course, whatever its name, the $\otimes$ operator continues to be useful to statisticians as in Vartak (1955), for example; also, in calculating Jacobians in multivariate distributions, e.g., Henderson and Searle (1979).

A still current use for determinants, at least in some teaching of elementary econometrics, is the solving of linear equations by the somewhat outdated method of Cramer (1750). This solves $\mathbf{Ax} = \mathbf{y}$ of $\mathbf{x}$ for the i'th element as $|\mathbf{A}_i^*|/|\mathbf{A}|$ where $\mathbf{A}_i^*$ is $\mathbf{A}$ with its i'th column replaced by $\mathbf{y}$. And, of course, a basic knowledge of determinants enables one to understand the elements of the inverse of a non-singular matrix.

3 MATRICES ENTER STATISTICS IN THE 1930s

With Gratton-Guiness (1994) having concluded (as previously quoted) that matrices "became prominent only from the 1920s", the year 1930 seems 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". Today that would undoubtedly be a welter of matrices; in 1930 the normal equations were solved using determinants. But two years later came the Turnbull and Aitken (1932) book with several applications of matrices to statistics, all of which are still important and widely used today: normal equations $\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$, and $\mathbf{X}'\mathbf{BX} = \mathbf{X}'\mathbf{By}$ if errors are assumed to be *not* of "equal importance and uncorrelated"; and $\mathrm{E}(\mathbf{x}'\mathbf{Ax}) = \mathrm{tr}(\mathbf{AV})$ for $\mathbf{x} \sim (\mathbf{0}, \mathbf{V})$

obtained by what today is considered to be a very roundabout derivation.

Then comes the Hotelling (1933) paper on principal components, a landmark both for its content and its matrix usage, although it does make reference to Kowalewski (1909) on determinants. Cochran (1934) is a next important milestone, dealing with the distribution of quadratic forms, and in the same year in Bartlett (1934) we find an early use in Britain of the word vector in a title—but no matrices in the paper. And the absence of matrices in Kolodziejczyk (1935) and in Welch (1935) is glaringly noticeable on today's standards; this is also true of Cochran (1938). All three of these papers deal with topics which lend themselves so easily to matrix representation, namely linear hypotheses and regression. But even the popular layman's book of that time, *Mathematics for the Millions*, Hogben (1936), has no index entry for determinants and only a single reference to matrix algebra, as being but one of several names used for "different ways of counting and measuring"! Contrarywise, Frazer, Duncan and Collar (1938) is a substantive book on matrices with virtually nothing in statistics but many applications in engineering, such as the oscillations of a triple pendulum (p. 310) and the disturbed steady motion of an aeroplane (p. 284). So the progress of matrices was slow. Infusion *had* begun, but the steeping was taking a long time. Nevertheless, it was to gain speed.

Craig (1938) as a follow-on from Cochran (1934), was the first in a long line of papers (by many authors) continuing to the present day, dealing with the independence of quadratic forms of normally distributed variables (see Section 4.1). Aitken (1937, 1938) continued publishing matrix results that would be of later value to statistics but with no explicit mention of statistics. The Girshick (1939) and Lukomski (1939) papers used determinants, Yates and Hale (1939) could probably have used matrices (but did not) in their discussion of missing rows, columns or treatments from Latin squares; and Bishop (1939) used a determinant as a generalized variance but has nothing on matrices. But the year 1939 ended with a flourish: the publication of Aitken's two books referred to at the beginning of this paper.

4 SPECIAL TOPICS

This section briefly describes some of the progress that matrices have made into certain statistical topics. The choice of topics is undoubtedly personal and the selection of references is certainly not complete.

4.1 Quadratic forms

Turnbull and Aitken (1932) show how to use a canonical form of a (symmetric) matrix to reduce a quadratic form to a sum of squares. Later, as already mentioned, Craig (1938) was an initial paper on the independence of quadratic forms (of "certain estimates of variance", actually), followed by Craig (1943) on "certain quadratic forms". In between, Hsu (1940) touched on the subject, Aitken (1940) dealt with the independence of linear and quadratic forms, and a decade later, Aitken (1950) dealt with the independence of two quadratic forms. Marten (1949) also contributes. The necessity condition for independence is still a hot topic; Searle (1971) gets it wrong. Graybill (1976) avoids the issue, Driscoll *et al.* (1986, 1988, 1995) have several proofs, Harville (1997) has a proof and Styan (1998) is known to be working on yet another proof. Lancaster (1956) confines attention to traces and cumulants.

4.2 Multivariate statistics

This branch of statistics has spawned more matrix activity than any other except, perhaps, for linear models.

Wilks (1932) uses lots of determinants in his work on the multiple correlation coefficient; but he uses almost no matrices and certainly no matrix algebra. In contrast, Lederman (1940) deals with a matrix problem arising from factor analysis. And Bartlett (1941) and Hsu (1941) in their discourses on canonical correlations use matrices aplenty. Bartky (1943) uses matrices sparingly, but including $\mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \cdots = (\mathbf{I} - \mathbf{M})^{-1}$ without proof or a reference thereto. Bartlett (1947) makes considerable use of matrices in his long "Multivariate Statistics" paper wherein he writes that he has "avoided complicated analytical discussion of theory" but has made use of "matrix and vector algebra". In doing

so, there is reference to Bartlett (1934) and to a paper by Tukey, "Vector methods in analysis of variance", described as having been on the programme at Princeton, November 1, 1946.

From this time on, matrices gather momentum towards becoming standard notation for multivariate statistics, peaking in the classic book by Anderson (1958), and continuing through to the present day in numerous books and papers, including the start of the *Journal of Multivariate Analysis* in 1971.

A matrix operation of particular use in multivariate statistics is the vectorizing of a matrix: writing its columns one under the other in one long vector. Its origin goes back to a Cambridge and London contemporary of Cayley, namely Sylvester (1884), and it is now usually called vec. It, and vech, an adaptation of vec to symmetric matrices developed in Searle (1978), are considered at length in Henderson and Searle (1979, 1981). A recent use is Wong and Li (1997).

Two important features of multivariate analysis are eigenroots and eigenvectors (known originally as latent roots and latent vectors—see Farebrother, 1997, p. 9, for interesting comment). Whatever their name they are matrix characteristics which feature frequently in multivariate literature. Early examples are Hsu (1940) on analysis of variance, Anderson (1948), Geary (1948) and Whittle (1953) in dealing with time series, with continuing interest through to Mallows (1961), concerned with "latent vectors of random symmetric matrices", and Shi (1997) having a section on "perturbation theory of eigenvalues and eigenvectors of a real symmetric matrix."

4.3 Solving normal equations of full rank

Normal equations derived from applying least squares to multiple regression data are usually of full rank. According to Kruskal and Stigler (1997, pp. 91-2) they have had that name since Gauss (1822). Their derivation and numerical solution have always been a source of great concern. Some early publications presenting normal equations are Wicksell (1930), Aitken and Silverstone (1942) and Bacon (1938) —all of whom use no matrices; but of course Turnbull and Aitken (1932) do. Until the acceptance of matrix inverses, methods for solving normal equations, being, as they are, just simultaneous linear equations, were

simply successive elimination with back substitution—or the Cramer's (1750) method using determinants as described previously.

Despite its algebraic clarity, the use of a matrix inverse initially posed considerable arithmetic difficulty. Although today's computers now make light work of that arithmetic, that has come about in only the last forty years; and it is accomplished at speeds that were utterly unimaginable then. During graduate student days in a small computing group at Cornell, there was great excitement when in 1959 we inverted a 10-by-10 matrix in seven minutes. After all, only a year or two earlier a friend had inverted a 40-by-40, by hand, using electric (Marchant or Monroe) calculators. That took six weeks! So it is understandable that statisticians were interested in computational techniques for inverting matrices—and they still are, for that matter, although at a much more sophisticated level than in the pre-computer days. An early beginning to those days was the Doolittle system for doing, and setting out, the individual calculations. The date and location of the publication of this system is perhaps surprising: Doolittle (1878) in the *U.S. Coast and Geodetic Report (1878)*, pages 115-120. Numerous abbreviations, improvements and comments followed, in a variety of publications, including in the statistical literature Horst (1938) and Dwyer (1941a, 1941b, 1944). In his 1944 paper, Dwyer wrote "The reader should be familiar with elementary matrix theory such as that outlined on pages 1-57 of Aitken's book" (1939a). An alternative approach was Bingham (1941) using the Cayley-Hamilton theorem and then Newton's equations. Other contributions included Hotelling (1943), Quenouille (1940), Fox (1950) and Fox and Hayes (1951). The arrival of computers soon put paid to these pencil-and-paper based methods, which by then had collected other variants of Doolittle such as abbreviated Doolittle, Crout-Doolittle, and so on.

4.4 Design of experiments

In the 1950s a widely used-book on experiment design was Cochran and Cox (1950, 2nd ed.); it has but one, brief, mention of matrices, related to regression. But only a year later Box and Wilson (1951) make substantial use of matrices for normal equations, their solution and resulting sums of squares. Likewise, Tocher (1952) has solid matrix

usage. Both papers give, for order n, $(\mathbf{I} + a\mathbf{J})^{-1} = \mathbf{I} - a\mathbf{J}/(1 + an)$; and Tocher extends this to $(\mathbf{I} + \mathbf{A} \otimes \mathbf{J})^{-1} = \mathbf{I} - [(\mathbf{I} + n\mathbf{A})^{-1}\mathbf{A}] \otimes \mathbf{J}$. This is an example of a class of designs involving its own special matrices. There are many other examples, one of the most extensive being factorial designs, for which Cornfield and Tukey (1956) developed far-reaching use of Kronecker (or direct) products of matrices for fixed effects models. Searle and Henderson (1979) extended this to the dispersion matrix for models involving random effects. Another example is Latin squares, for which Yates and Hale (1939) could have benefited from using matrices—but they did not. But Cox (1956) certainly did—in abundance. Design features involving circulant matrices include autocorrelation and spectral density functions (Wise, 1955), circular stationary models (Olkin and Press, 1969) and partial factorial balance (Anderson, 1972); for these Searle (1979) gives methods for inverting circulants of two and three non-zero elements per row. Block designs have always been a fertile field for the use of matrices (although, of course, Fisher made no use of them), two papers in the 1980s being John and Williams (1982) on optimal designs and Constantine (1983) on balanced incomplete block designs. And the use of matrices continues apace in this area.

4.5 Linear models

H.O. Hartley makes an interesting comment in the published discussion of Tocher (1952, p.96), suggesting that it was Barnard who introduced the unifying principle of developing "analysis of variance by analogy to regression analysis". This is, of course, the vital connection for what used to be called "fitting constants" and is now usually known as linear models. It is the foundation on which enormous numbers of research papers have been built—and also, numerous books, so many in fact that commenting on them is well beyond the scope of this paper. And certainly during the last forty or more years, all of them use matrices as the *lingua franca*. The earliest book to do so seems to be Kempthorne (1952). Nevertheless, seven years later the Williams' (1959) book on regression had only a tiny mention of matrices.

One of the greatest contributions to understanding the apparent quirkiness of normal equations of non-full rank (as is customary with linear models), which have an infinity of

solutions, is due to Rao (1962). Using the work of Moore (1920) and Penrose (1955), he showed how a generalized inverse matrix yields a solution to the normal equations, and how that solution can be used to establish estimable functions and their estimators—and these results are invariant to whatever generalized inverse is being used. Although the arithmetic of generalized inverses is scarcely any less than that of regular inverses, the use of generalized inverses is of enormous help in understanding estimability and its consequences.

4.6 Other topics

Of the many other topics and places where matrices have contributed substantially to statistics, a few are mentioned briefly here.

4.6.1 Probability theory and Markov Chains

Back-to-back papers using matrices for branching processes and queuing processes are Hammersley and Morton (1954) and Bailey (1954). More recently, generalized inverse matrices (so useful to linear models) have been incorporated by Hunter (1982, 1988 and 1990) into applications of applied probability and of Markovian kernels. They also play an important role in his book, Hunter (1983), its fourth chapter, "Matrix Techniques", constituting some 20% of the book.

4.6.2 Age distribution vectors

When the proportions (or numbers) in different age groups of an animal population are arrayed in a vector it is an age distribution vector, Lewis (1942). That vector is seldom constant over time. Leslie (1945, 1948), by using age-specific birth rates as the first row of a matrix, and age-specific survival rates as the sub-diagonal of that matrix (now called a Leslie matrix) shows how pre-multiplying an age-distribution vector for time t by a Leslie matrix yields (deterministically) the age distribution vector at time $t + 1$. In a practical application of this, using monthly Leslie matrices and a killing (diagonal) matrix, Darwin and Williams (1964) adapt this procedure to ascertain optimum months of the year for poisoning rabbits on New Zealand farms. Rabbits there are a seriously overpopulated pest,

consuming grass needed for producing meat, wool and milk which sustain the country's economy through exports.

4.6.3 Random effects in linear models

In regression, and in the earliest use of linear models for "fitting constants", the elements of $\boldsymbol{\beta}$ in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ were deemed to be constants, or fixed effects. But in the 1940s geneticists, in particular, wanted to include in $\boldsymbol{\beta}$ random variables (or more truthfully, realized but unobservable values of random variables) representing genetic worth of animals whose production records were to be analyzed. These soon came to be called random effects. Incorporating them into $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ solely by having var(e) be something other than $\sigma^2\mathbf{I}$ (to include the variance structure of the random effects) made it impossibly difficult to estimate both the fixed effects and the variances of the random effects. Thus was born the representation

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \tag{1}$$

where $\boldsymbol{\beta}$ contains the fixed effects and $\mathbf{u}$ the random effects, with $\mathbf{X}$ and $\mathbf{Z}$ being the corresponding incidence matrices. This simple, matrix motivated rewriting of the model has been of tremendous value, and is the basis for a vast array of papers and book chapters.

A particularly valuable extension of $\mathbf{Zu}$ in (1) is that of $\mathbf{Zu} = \Sigma\mathbf{Z}_i\mathbf{u}_i$ where each $\mathbf{u}_i$ has all the random effects occurring in the data of a single (main effect or interaction) factor. This was initiated by Hartley and Rao (1967) and with $\mathrm{var}(\mathbf{u}_i) = \sigma_i^2\mathbf{I}$ and $\mathrm{cov}(\mathbf{u}_i, \mathbf{u}_t') = \mathbf{0}$ for $i \neq t$, it enabled them to derive equations for maximum likelihood (under normality) estimation of both $\boldsymbol{\beta}$ and the variances σ_i^2. This was a major breakthrough in the problem of estimating variance components, and has become a standard estimation procedure.

4.6.4 Estimating genetic worth

Programs for improving the genetic worth of an animal species through planned matings often include estimating the genetic worth of an untried animal from records of its ancestors. Searle (1963), shows that the correlation between genetic worth and an ancestor-based estimate of it is of the form $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$ where $\mathbf{x}'$ is closely related to rows of $\mathbf{A}$. As a result,

through knowing how elements of $\mathbf{A}^{-1}$ involve cofactors and the determinant of $\mathbf{A}$, a recurrence relationship for the correlation (based on the number of generations of ancestors) was established. Conclusions are that going beyond grandparents contributes very little to one's estimate of the untried animal. (Horse-racing enthusiasts wanting to buy an untried yearling should take heed!)

4.6.5 Inverting $\mathbf{A} + \mathbf{UBV}$

The matrix $\mathbf{A} + \mathbf{UBV}$ arises in various forms in statistics. For example, one special case is the dispersion matrix $\mathbf{D}(\mathbf{p}) - \mathbf{pp}'$ for a multinomial random variable, where $\mathbf{p}$ is a vector of probabilities and $\mathbf{D}(\mathbf{p})$ is the diagonal matrix of elements of $\mathbf{p}$. It also occurs as an intraclass correlation matrix $(1-\rho)\mathbf{I}+\rho\mathbf{J}$, it arises as discriminant analysis, Bartlett (1951), and it turns up when elements of a matrix are altered, Sherman and Morrison (1950).

What is interesting about the inverse of $\mathbf{A} + \mathbf{UBV}$ is that although

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1},$$

as the standard result is widely known, there are many variations of it, as well as numerous special cases. Henderson and Searle (1981) give a number of these as well as much of the history. They and Harville (1997) also deal with generalized inverses.

4.6.6 Partitioned inverses

The value of generalized inverses in linear models analysis has already been mentioned. They are particularly useful when dealing with partitioned models $\mathrm{E}(\mathbf{y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$. For

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^* = \begin{bmatrix} \mathbf{A}^- & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^-\mathbf{B} \\ \mathbf{I} \end{bmatrix} (\mathbf{D} - \mathbf{CA}^-\mathbf{B})^-[-\mathbf{CA}^- \quad \mathbf{I}]. \tag{2}$$

$\mathbf{Q}^*$ is a generalized inverse of $\mathbf{Q}$ provided, for $r(\mathbf{Q})$ being the rank of $\mathbf{Q}$,

$$r(\mathbf{Q}) = r(\mathbf{A}) + r(\mathbf{D} - \mathbf{CA}^-\mathbf{B}), \tag{3}$$

as in Marsaglia and Styan (1974). For non-singular $\mathbf{Q}$ the generalized inverse $\mathbf{Q}^-$ becomes the regular inverse $\mathbf{Q}^{-1}$. In the case of $\mathbf{Q} = \mathbf{X}'\mathbf{X}$ for $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$, the rank condition

(3) always holds, and $\mathbf{Q}^* = \mathbf{Q}^-$. The use of $\mathbf{Q}^-$ in providing $\tilde{\boldsymbol{\beta}} = \mathbf{Q}^-\mathbf{X}'\mathbf{y}$ then leads immediately to very useful results such as $\tilde{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_2$ and $\mathrm{R}(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$, the reduction in sum of squares for fitting $\boldsymbol{\beta}_1$ adjusted for $\boldsymbol{\beta}_2$.

5 Books

Tracing the intermingling of matrices and statistics in published books could be a long process. It will here be brief, in terms of a dichotomy: the occurrence of matrices in statistics books, and of statistics in matrix books.

Fisher (1935 *et seq*) had no matrices in any of its numerous editions. Many sets of numbers are laid out in rectangular arrays, but they are designs, not matrices. Snedecor (1937), Kendall (1943-1952), Mood (1950), and Mood and Graybill (1963) had either no matrices or almost none, but by Kendall and Stuart (1958, Vol. I, 1st Ed.), their Chapters 15 and 19 had matrix notation for the multi-normal distribution, for quadratic forms, and for least squares. But the Snedecor book took much longer to join the matrix crowd. It is in Snedecor and Cochran (1989, 8th Ed.), where the preface heralds the arrival of matrices with the following near-apologia:

> " A significant change in this edition occurs in the notation used to describe the operations of multiple regression. Matrix algebra replaces the original summation operators, and a short appendix on matrix algebra is included."

Cramér (1946) was early in having a whole chapter "Matrices, determinants and quadratic forms", referencing Cochran (1934) for the last of those three topics. But Kempthorne (1952) and Rao (1952) seem to be the first books having substantial sections in matrix notation and making considerable use of matrix algebra. Rao (1952) begins with some thirty pages of matrices, having written in his preface "The problems of multivariate analysis resolve themselves into an analysis of the dispersion matrix and reduction of determinants." Thereafter came Anderson (1958), Rao (1965), Searle (1971), Graybill (1976), Seber (1977), Hocking (1997) and many others too numerous to mention, all using matrices extensively. So do Kotz *et al.* (1985) in the *Encyclopedia of Statistical Sciences* (Vol. 5) and the *Ency-*

clopedia of Biostatistics, the latter having "Matrix Algebra" as a major entry.

Now to the occurrence of statistics in matrix books. Those by Frazer *et al.* (1938), Aitken (1939a) and Ferrar (1941) have no statistics. Bellman (1960) has two chapters (10%) on Markov matrices and probability theory. And the Graybill (1969, 1983) books have plenty of statistics, as one would expect from their titles. The same is true of Searle (1982) and Magnus and Neudecker (1988), also of Harville (1997) which is very comprehensive, and of Schott (1997) which is at a more elementary level with but a modest amount of statistics.

Then there are the specialized books on generalized inverses, three in the same year: Pringle and Rayner (1971), Rao and Mitra (1971), both with plenty of statistics, and Boullion and Odell (1971) with a modest tilt to statistics. Shortly thereafter came Ben-Israel and Greville (1974) with virtually none. There are also specialty books on matrix calculus, Rogers (1980) with little statistics, and Magnus and Neudecker (1988) with plenty.

Whilst this incursion of statistics in matrix books was taking place, we can note that it was also happening in linear algebra journals. In *Linear Algebra and its Applications* (begun in 1968 and then referred to as LAA) and *Linear and Multilinear Algebra* (started 1972) the early issues showed little evidence of statistics, whereas nowadays there is quite a good representation of statistics. In particular, LAA periodically now has special issues on matrices and statistics; the 1990 such issue had 650 pages. Many papers in these issues stem from the new almost annual meeting of the Workshop on Matrices and Statistics. Started in 1990, the only years of this decade in which the workshop has not met are 1991 and 1993. So matrices in statistics are alive and well.

OMISSIONS

To those whose work has not been mentioned, my apologies. For a hint at the vastness (in both time and geography) of the literature, they are referred to Puntanen and Styan (1988). It has 1,596 literature references on matrices and statistics; and many many more have been added in supplements during the last eleven years.

References

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

Aitken, A.C. (1940) On the independence of linear and quadratic forms in samples of normally distributed variates. *Proceedings, Royal Society of Edinburgh* **60**, 40-46.

Aitken, A.C. (1939a) *Determinants and Matrices*, Oliver & Boyd, Edinburgh.

Aitken, A.C. (1939b) *Statistical Mathematics*, Oliver & Boyd, Edinburgh.

Aitken, A.C. (1938) Studies in practical mathematics. II. The evaluation of the latent roots and latent vectors of a matrix. *Proceedings, Royal Society of Edinburgh*, **LVII**, 269-304.

Aitken, A.C. (1937) Studies in practical mathematics. I. The evaluation, with applications, of a certain triple product matrix. *Proceedings, Royal Society of Edinburgh*, **LVII**, 172-181.

Aitken, A.C. and Silverstone, H. (1942) On the estimation of statistical parameters. *Proceedings, Royal Society of Edinburgh, Series A* **61**, 186-194.

Anderson, D.A. (1972) Designs with partial factorial balance. *Annals of Mathematical Statistics*, **43**, 1333-1341.

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

Anderson, T.W. (1948) The asymptotic distribution of the roots of certain determinantal equations. *Journal of the Royal Statistical Society (B)*, **X**, 132-139.

Bacon, H.M. (1938) Note on a formula for the multiple correlation coefficient. *Annals of Mathematical Statistics* **9**, 227-229.

Bailey, N.T.J. (1954) On queueing processes with bulk services. *Journal of the Royal Statistical Society (B)* **XVI**, 80-87.

Barkty, W. (1943) Multiple sampling, with constant probability. *Annals of Mathematical Statistics* **14**, 363-377.

Bartlett, M.S. (1951) An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics* **22**, 107-111.

Bartlett, M.S. (1947) Multivariate analysis. *Journal of the Royal Statistical Society (B)* **IX**, 176-197.

Bartlett, M.S. (1941) The statistical significance of canonical correlations. *Biometrika* **32**, 29-37.

Bartlett, M.S. (1934) The vector representation of a sample. *Proceedings, Cambridge Philosophical Society* **30**, 327-340.

Bell, E.T. (1953) *Men of Mathematics*, Vol. 2, Penguin, London, 441-443.

Bellman, R. (1960) *Introduction to Matrix Analysis*, McGraw-Hill, New York.

Ben-Israel, A. and Greville. N.E. (1974) *Generalized Inverses: Theory and Applications*, Wiley, New York.

Bingham, H.D. (1941) A new method for obtaining the inverse matrix. *Journal of the American Statistical Association* **36**, 530-534.

Bishop, D.J. (1939) On a comprehensive test for the homogeneity of variances and covariances in multivariate problems. *Biometrika* **31**, 31-55.

Boullion, T.L. and Odell, P.L. (1971) *Generalized Inverse Matrices*, Wiley, New York.

Box, G.E.P. and Wilson, K.B. (1951) On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society* **XIII**, 1-45.

Cayley, A. (1858) A memoir on the theory of matrices. *Philosophical Transactions of the Royal Society of London* **148**, 17-37.

Cayley, A. (1855) Sept differents memoires d'analyse. No. 3. Remarqu sur la notation des functions algebriques. *Journal für die reine und angewandte Mathematik* **50**, 282-285.

Cochran, W.G. (1938) The omission or addition of an independent variate in multiple linear regression. *Journal of the Royal Statistical Society (B)*, **V**, 171-176.

Cochran, W.G. (1934) The distribution of quadratic forms in a normal system, with applications to analysis of covariance. *Proceedings, Cambridge Philosophical Society* **30**, 178-191.

Cochran, W.G. and Cox, G.M. (1950) *Experimental Designs*, Wiley, New York.

Constantine, G.M. (1983) On the trace efficiency for control of reinforced balanced incomplete block designs. *Journal of the Royal Statistical Society (B)* **45**, 31-36.

Cornfield, J. and Tukey, J.W. (1956). Average values of mean squares in factorials. *Annals of Mathematical Statistics* **27**, 907-949.

Cox, C.P. (1956) The analysis of Latin square designs with individual curvature in one direction. *Journal of the Royal Statistical Society (B)* **XX**, 193-204.

Craig, A.T. (1943) Note on the independence of certain quadratic forms. *Annals of Mathematical Statistics* **14**, 195-97.

Craig, A.T. (1938) On the independence of certain estimates of variance. *Annals of Mathematical Statistics* **9**, 48-55.

Cramer, G. (1750) *Introduction à l'Analyse des Lignes Courbe Algébriques*, Cramer et Philibert, Geneva.

Cramèr, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey.

Cullis, C.E. (1913-25) *Matrices and Determinoids*, 3 volumes, Cambridge University Press, Cambridge.

de la Valèe Pousain, C.J. (1911) Sur la méthode de l'approximation minimum. *Annales de la Société Scientifique de Bruxelles* **35**, 1-16.

Darwin, J.H. and Williams, R.M. (1964) The effect of time of hunting on the size of a rabbit population. *New Zealand Journal of Science* **7**, 341-352.

Doolittle, M.H. (1878) Method employed in the solution of normal equations and the adjustment of a triangulation. *U.S. Coast and Geodetic Survey Report (1878)*, 115-120.

Driscoll, M.F. and Gundberg, W.R. (1986) The history of the development of Craig's theorem. *The American Statistician* **40**, 65-71.

Driscoll, M.F. and Krasnicka, B. (1995) An accessible proof of Craig's theorem in the general case. *The American Statistician* **49**, 59-61.

Driscoll, M.F. and Reid, J.G. (1988): see Reid and Drisoll (1988).

Dwyer, P.S. (1941a) The Doolittle technique. *Annals of Mathematical Statistics* **12**, 449-458.

Dwyer, P.S. (1941b) The solution of simultaneous linear equations. *Psychometrika* **6**, 101-129.

Dwyer, P.S. (1944) A matrix presentation of least squares and correlation theory with matrix justification of improved methods of solution. *Annals of Mathematical Statistics* **15**, 82-89.

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

Farebrother, R.W. (1997) A.C. Aitken and the consolidation of matrix theory. *Linear Algebra and its Applications* **43**, 3-12.

Farkas, J. (1902) Theorie der einfachen Ungleichungen. *Journal für die Reine und Angewandte Mathematik* **124**, 1-27.

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

Fisher, R.A. (1935) *The Design of Experiments*, Oliver & Boyd, London.

Fox, L. (1950) Practical methods for the solution of linear equations and the inversion of matrices. *Journal of the Royal Statistical Society (B)* **XII**, 120-136.

Fox, L. and Hayes, J.G. (1951) More practical methods for the inversion of matrices. *Journal of the Royal Statistical Society (B)* **XIII**, 83-91.

Frazer, R.A., Duncan, W.J. and Collar, A.R. (1938) *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, Cambridge University Press, Cambridge.

Gauss, Herr Hofrath und Ritter [Carl Friedrich] (1822) Anwendung der Wahrscheinlichkeitsrechnung auf eine Aufgabe der practishen Geometrie. *Astronomische Nachrichten* 1/6 columns, 81-86.

Geary, R.C. (1948) Studies in relations between economic time series. *Journal of the Royal Statistical Society (B)* **X**, 140-158.

Girshick, M.A. (1939) On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics* **10**, 203-224.

Grattan-Guiness, I. (1994) A new type of question: on the pre-history of linear and non-linear programming, 1770-1940. In *The History of Modern Mathematics* **3**, 43-89.

Grattan-Guiness, I. and Lederman, W. (1994) Matrix theory. In *Companion Encyclopedia to the History and Philosophy of the Mathematical Sciences* (I. Gratton-Guiness, Ed.), Routledge, London, 775-786.

Graybill, F.A. (1983) *Matrices with Applications in Statistics*, Wadsworth, Belmont, California.

Graybill, F.A. (1976) *Theory and Application of the Linear Model*, Duxbury Press, North Scituate.

Graybill, F.A. (1969) *Introduction to Matrices with Applications in Statistics*, Wadsworth, Belmont, California.

Haar, A. (1924) Uber linear Ungleichungen. *Acta litterarum ac scientarium regale Universitatis Hungaricae Francisco-Josephine. sectio scientarum mathematicum* **2**, 1-14.

Hammersley and K.W. Morton, K.W. (1954) Transposed branching processes. *Journal of the Royal Statistical Society (B)* **XVI**, 76-79.

Hartley, H.O. and Rao, J.N.K. (1967) Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93-108.

Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective*, Springer, New York.

Henderson, H.V., Pukelsheim, F. and Searle, S.R. (1983) On the history of the Kronecker product. *Linear and Multilinear Algebra* **14**, 113-120.

Henderson, H.V. and Searle, S.R. (1981a) The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra* **9**, 271-288.

Henderson, H.V. and Searle (1981b) On deriving the inverse of a sum of matrices. *SIAM Review* **23**, 53-60.

Henderson, H.V. and Searle, S.R. (1979) Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics* .4in **7**, 65-81.

Hensel, K. (1891) Uber die Darstellung der Determinante eines Systems, welches aus zwei anderen componint ist. *Acta Mathematica* **14**, 317-319.

Hocking, R.R. (1997) *Methods and Applications of Linear Models*, Wiley, New York.

Hogben, L. (1936) *Mathematics for the Millions*, Allen & Unwin Ltd., London.

Horst, P. (1932) A short method for solving for a coefficient of correlation. *Annals of Mathematical Statistics*, **3**, 40-44.

Hotelling, H. (1943) Some new methods in matrix calculation. *Annals of Mathematical Statistics* **14**, 1-34.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology* **24**, 417-441.

Hsu, P.L. (1941) On the limiting distribution of the canonical correlations. *Biometrika* **32**, 38-45.

Hsu, P.L. (1940) On generalized analysis of variance. *Biometrika* **31**, 221-237.

Hunter, J.J. (1983) *Mathematical Techniques of Applied Probability: Vol. 1, Discrete Time Models: Basic Theory*, Academic Press, New York.

Hunter, J.J. (1990) Parametric forms for generalized inverses of Markovian kernels and their applications. *Linear Algebra and its Applications* **127**, 71-84.

Hunter, J.J. (1988) Characterizations of generalized inverses associated with Markovian kernels. *Linear Algebra and its Applications* **102**, 121-142.

Hunter, J.J. (1982) Generalized inverses and their application to applied probability problems. *Linear Algebra and its Applications* **45**, 157-198.

John, J.A. and Williams, E.R. (1982) Conjectures on optimal block designs. *Journal of the Royal Statistical Society (B)* **44**, 221-225.

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

Kendall, M.G. (1943) *The Advanced Theory of Statistics*, Griffin, London.

Kolodziejczyk, S. (1935) 'On an important class of statistical hypotheses. *Biometrika* **27**, 161-190.

Kotz, S., Johnson, N.L. and Read, C.B., Eds. (1985) *Encyclopedia of Statistical Sciences*, Vol. 5, Wiley, New York.

Kowalewski, G. (1909) *Einführing in die determinantentheorie einschliesslich der unendlichen und der Fredholmschen determinanten.* Veit & comp., Leipzig.

Kruskal, W.H. and Stigler, S.M. (1997) *Statistics and Public Policy*, Ed., Bruce D. Spencer, Clarendon Press, Oxford.

Lancaster, H.O. (1954) Traces and cumulants of quadratic forms in normal variables. *Journal of the Royal Statistical Society* **16**, 247-254.

Lederman, W. (1940) On a problem concerning matrices with variable diagonal elements. *Proceedings, Royal Society of Edinburgh* **LX**, 1-17.

Leslie, P.H. (1948) Some further notes on the use of matrices in population mathematics. *Biometrika* **35**, 213-245.

Leslie, P.H. (1945) On the use of matrices in certain population mathematics. *Biometrika* **33**, 183-212.

Lewis, E.G. (1942) On the generation and growth of a population. *Sankhyā* **6**, 93-96.

Lukomski, J. (1939) On some properties of multidimensional distributions. *Annals of Mathematical Statistics* **10**, 236-246.

Magnus, J.P. and Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

Mallows, C.L. (1961) Latent vectors of random symmetric matrices. *Biometrika* **48**, 133-149.

Marsaglia, G. and Styan, G.P.H. (1974) Rank conditions for generalized inverses of partitioned matrices. *Sankhyā, Series A* **36**, 437-442.

Marten, B. (1949) Independence of non-negative quadratic forms in normally correlated variables. *Annals of Mathematical Statistics* **20**, 119-120.

Mood, A.M. (1950) *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Mood, A.M. and Graybill, F.A. (1963) *Introduction to the Theory of Statistics*, 2nd ed., McGraw-Hill, New York.

Moore, E.H. (1920) On the reciprocal of the general algebraic matrix. *Bulletin, American Mathematical Society* **26**, 394-395.

Muir, T. (1960, reprint of 1890, 1911, 1920, 1923) *The Theory of Determinants in the Historical Order of Development*, Dover, New York.

Olkin, I. and Press, S.S. (1969) Testing and estimation for a circular stationary model. *Annals of Mathematical Statistics* **40**, 1358-1373.

Penrose, R.A. (1955) A generalized inverse for matrices. *Proceedings, Cambridge Philosophical Society* **51**, 406-413.

Pringle, R.M. and Rayner, A.A. (1971) *Generalized Inverse Matrices with Applications to Statistics*, Griffin, London.

Puntanen, S. and Styan, G.P.H. (1988) *A Personal Guide to the Literature in Matrix Theory for Statistics and Some Related Topics*, Report A205, Department of Statistical Sciences, University of Tampere, Finland.

Quenouille, M.H. (1950) Computational devices in the application of least squares. *Journal of the Statistical Society (B)* **XII**, 258-272.

Rao, C.R. (1965) *Linear Statistical Inference and its Applications*, Wiley, New York.

Rao, C.R. (1962) A note on a generalized inverse of a matrix with applications to problems in statistics. *Journal of the Royal Statistical Society (B)* **24**, 152-158.

Rao, C.R. (1952, reprint 1970) *Advanced Statistical Methods in Biometric Research*, Hafner, Darien, Connecticut.

Rao, C.R. and Mitra, S.K. (1971) *Generalized Inverse of Matrices and its Applications*, Wiley, New York.

Reid, J.G. and Driscoll, M.F. (1988) An accessible proof of Craig's theorem in the noncentral case. *The American Statistician* **42**, 139-142.

Rogers, G.S. (1980) *Matrix Derivatives*, Marcel Dekker, New York.

Schott, J.R. (1997) *Matrix Analysis for Statistics*, Wiley, New York.

Searle, S.R. (1982) *Matrix Algebra Useful for Statistics*, Wiley, New York.

Searle, S.R. (1979) On inverting circulant matrices. *Linear Algebra and its Applications* **25**, 77-89.

Searle, S.R. (1978) A univariate formulation of the multivariate linear model. In *Contributions to Survey Sampling and Applied Statistics, Papers in Honor of H.O. Hartley* (H.A. David, ed.), Academic Press, New York, 181-189.

Searle, S.R. (1971) *Linear Models*, Wiley, New York.

Searle, S.R. (1963) The efficiency of ancestor records in animal selection. *Heredity* **18**, 351-360.

Searle, S.R. and Henderson, H.V. (1979) Dispersion matrices for variance components models. *Journal of the American Statistical Association* **74**, 465-470.

Seber, G.A.F. (1977) *Linear Regression Analysis*, Wiley, New York.

Sherman, J. and Morrison, W.J. (1950) Adjustment of an inverse matrix corresponding to a change in one element of a matrix. *Annals of Mathematical Statistics* **21**, 124-127.

Shi, L. (1997) Local influence in principal components analysis. *Biometrika* **84**, 175-186.

Snedecor, G.W. (1937-56) *Statistical Methods*, Editions 1-5, Iowa University Press, Ames, Iowa.

Snedecor, G.W. and Cochran, W.G. (1967-89) *Statistical Methods*, Editions 6-8, Iowa University Press, Ames, Iowa.

Styan, G.P.H. (1998) Personal communication.

Sylvester, J. (1884) Sur la résolution générale de l'équation linéaire en matrices d'un ordre quelconque. *C. R. Acadamie Scientifique Paris* **99**, 409-12 and 432-36.

Tocher, K.D. (1952) The design and analysis of block experiments. *Journal of the Royal Statistical Society (B)* **XIV**, 45-100.

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

Vartak, N.N. (1955) On an application of Kronecker product of matrices to statistical design. *Annals of Mathematical Statistics* **26**, 420-438.

Welch, B.A. (1935) Some problems in the analysis of regression among k samples of two variables. *Biometrika* **27**, 145-160.

Whittle, P. (1953) The analysis of multiple stationary time series. *Journal of the Royal Statistical Society (B)* **XV**, 125-139.

Wicksell, S.D. (1930) Remarks on regression. *Annals of Mathematical Statistics* **1**, 3-13.

Wilks, S.S. (1932) On the sampling distribution of the multiple correlation coefficient. *Annals of Mathematical Statistics* **3**, 196-203.

Williams, E.S. (1959) *Regression Analysis*, Wiley, New York.

Wise, J. (1955) The autocorrelation function and the spectral density function. *Biometrika* **42**, 151-159.

Wong, H. and Li, W.K. (1997) On a multivariate conditional heteroscedastic model. *Biometrika* **84**, 111-123.

Yates, F. and Hale, R.W. (1939) The analysis of Latin squares when two or more rows, columns or treatments are missing. *Journal of the Royal Statistical Society (B)* **VI**, 67-79.

Zehfuss, G. (1858) Uber eine gewisse Determinante. *Zeitschrift für Mathematik und Physik* **3**, 298-301.