A MATRIX PRESENTATION OF LEAST SQUARES AND CORRELATION THEORY WITH MATRIX JUSTIFICATION OF IMPROVED METHODS OF SOLUTION

By Paul S. Dwyer

University of Michigan

1. Introduction and summary. It is the aim of this paper to exhibit, by using elementary matrix theory, the basic concepts of least squares and correlation theory, the solution of the normal equations, and the presentation and justification of recently developed and newly proposed techniques into a single, compact, and short presentation. We shall be mainly concerned with the following topics:

a. Basic least squares theory including derivation of normal equations, the theoretical solution of these equations (regression coefficients), the standard errors of these solutions, and the standard error of estimate.

b. The more specific theory (correlation theory) resulting from applying the general least squares results to the standardized distributions.

c. A matrix presentation of the Doolittle solution.

d. A simple matrix justification of methods, previously presented, for getting least squares and multiple correlation constants from the entries of an abbreviated Doolittle solution.

e. A presentation of a more general theory which the matrix presentation reveals.

f. The outline of a "square root" method as an alternative to theDoolittle method.

The reader should be familiar with elementary matrix theory such as that outlined on pages 1–57 of Aitken's book [1].

No previous knowledge of the Doolittle technique is demanded although a familiarity with the notation and contents of two earlier papers [2], [3] is advised, particularly for those who are interested in the computational aspects.

The presentation here is theoretical and is not concerned with such computational topics as the number of decimal places required, etc. With reference to the number of places, the reader is referred to the recent paper of Professor Hotelling [4].

2. Notation. Let $[x'_{ij}]$ with $1 \leq i \leq N$ and $1 \leq j \leq n$ be the n by N matrix of observed variates of n "predicting variables" for N individuals with i indicating the individual and j the variable. Let $[y'_i]$ be the one by N column matrix of the observed variates of the "predicted" variable. Let the matrices of deviations from the variable means be indicated by $[x_{ij}] = X$ and $[y_i] = Y$. Then by the least squares hypothesis we are to find numbers $b_{y1\cdots}, b_{y2\cdots}, \cdots, b_{yn\cdots}$ such that

$$e_i = y_i - (x_{i1}b_{y1\cdots} + x_{i2}b_{y2\cdots} + \cdots + x_{in}b_{yn\cdots}),$$

shall have a minimum variation (standard deviation). We then denote the $b_{yi\ldots}$ by the one by n column matrix B and the e_i by the one by N column matrix E and have

$$E = Y - XB, \tag{1}$$

as the basic matrix equation.

It may be noted further that the fitted values of y_i are given by the one by N matrix product $XB = \underset{\sim}{Y}$. Using this notation (1) appears as

$$E = Y - \underset{\sim}{Y}. \tag{1'}$$

3. Basic least squares theory.

a. Sum of squares of residuals. The condition for minimum variation in this situation (variates measured from means) is equivalent to the condition for minimum sum of squares of residuals. In matrix notation this sum of the squares of the residuals can be written

$$E'E \text{ with } E = Y - XB = Y - \underset{\sim}{Y}, \quad \text{and } E' \text{ the transpose of } E. \tag{2}$$

b. The normal equations. Differentiating (2) with respect to B' we find the necessary condition to be

$$X'E = 0. \tag{3}$$

This matrix equation gives the normal equations in implicit form. More explicitly by (1) we have $X'(Y - XB) = 0$ so

$$X'XB = X'Y. \tag{4}$$

The reader should immediately recognize that (4) is the matrix equivalent of the usual statement of the normal equations where deviations from the means are used. It should be noted also that (3) and (4) can be written in the form

$$X'\underset{\sim}{Y} = X'Y \quad \text{from whence at once } \underset{\sim}{Y}'\underset{\sim}{Y} = \underset{\sim}{Y}'Y. \tag{5}$$

c. Solution of normal equations. The theoretical solution of (4) is accomplished at once and results in

$$B = (X'X)^{-1}X'Y = (X'X)^{-1}X'\underset{\sim}{Y}. \tag{6}$$

d. Standard deviation of residuals. The standard deviation of residuals is $\sqrt{E'E/N}$.

In order to evaluate this we note that

$$\underset{\sim}{Y}'E = B'X'E = 0, \quad \text{and } \underset{\sim}{Y}'Y = \underset{\sim}{Y}'\underset{\sim}{Y}, \tag{7}$$

Thus

$$E'E = (Y - \underset{\sim}{Y})'E = Y'E = Y'(Y - XB) = Y'Y - Y'XB, \tag{8}$$

and

$$Y'XB = Y'\underset{\sim}{Y} = \underset{\sim}{Y}'\underset{\sim}{Y} = \underset{\sim}{Y}'Y. \tag{9}$$

Since $Y'Y = \Sigma y^2$, we have

$$E'E = \Sigma y^2 \left(1 - \frac{Y'XB}{\Sigma y^2}\right), \tag{10}$$

so that, dividing by N and taking the square root

$$s_e = s_y \sqrt{1 - \frac{Y'XB}{\Sigma y^2}}. \tag{11}$$

If the relation between the estimated standard deviations in the population is desired then divide each side of (10) by the number of degrees of freedom and get

$$\sigma_e = \sigma_y \sqrt{1 - \frac{Y'XB}{\Sigma y^2}}. \tag{12}$$

Alternative formulas to (11) and (12) are obtained by replacing $Y'XB$ by its equivalent expressions in (9).

e. Formulas for multiple correlation coefficient. It is to be noted that the numerical quantity $Y'XB/\Sigma y^2$ plays an important role in measuring the ratio σ_e/σ_y. It is customary to use this quantity as the definition of the square of the multiple correlation coefficient so we have

$$r^2_{y \cdot x_1 \cdots x_n} = \frac{Y'XB}{\Sigma y^2} = \frac{B'X'XB}{\Sigma y^2} = \frac{Y'X(X'X)^{-1}X'Y}{\Sigma y^2} = \frac{\underline{Y}'\underline{Y}}{\Sigma y^2} = \frac{\sigma^2_{\underline{y}}}{\sigma^2_y}. \tag{13}$$

f. Formulas for correlation coefficient. When $n = 1$, $X'X = \Sigma x^2$, $Y'X = X'Y = \Sigma XY$, $B = b$ and (13) gives

$$r_{yx} = \sqrt{\frac{b\Sigma xy}{\Sigma y^2}} = b\sqrt{\frac{\Sigma x^2}{\Sigma y^2}} \left(= b\,\frac{\sigma_x}{\sigma_y}\right) = \frac{\Sigma xy}{\sqrt{\Sigma x^2\, \Sigma y^2}} = \frac{\sigma_{\underline{y}}}{\sigma_y}. \tag{14}$$

Many of the above developments can be duplicated, without formal use of matrix theory, by judicious use of symbolism and substitution. See for example the presentations of Kirkham [5], Bacon [6], and Guttman [7].

g. Errors of regression coefficients. If B_0 is an approximation to B such that $B_0 + \Delta B = B$ then (6) can be written

$$B_0 + \Delta B = (X'X)^{-1}X'Y$$

and

$$\Delta B = (X'X)^{-1}X'(Y - XB_0). \tag{15}$$

This formula can be used in finding corrections ΔB necessary to change any proposed trial solution, B_0, into a correct solution. It could also be used in extending the accuracy of a solution after an approximation had been secured to a specific number of places. It has greater utility however in another problem.

We suppose that the predicting variables, the x's, contain no errors but that there are errors in the observed values of y. Let the hypothetical observed values of y be indicated by Y and the recorded observed values of y by Y_0. Let the values of B_0 be the regression coefficients obtained by using the recorded observed values Y_0. Thus $\Delta B = 0$ when Y is replaced by Y_0 in (15). Now let $Y - XB_0$, the "true" residual errors of the recorded observed values, be indicated by E. Then (15) becomes

$$\Delta B = (X'X)^{-1}X'E. \tag{16}$$

Sampling theory can be applied to (16) to obtain a formula for the standard error of the regression coefficient. It is assumed that the "true" residual errors are independent with a common standard deviation σ_e. The values of ΔB are then linear functions of these errors. It follows that

$$\sigma_B^2 = \sigma_{\Delta B}^2 = (X'X)^{-1}X'X(X'X)^{-1}\sigma_e^2 = (X'X)^{-1}\sigma_e^2. \tag{17}$$

The standard errors of the regression coefficients are thus formed by multiplying σ_e by the square roots of the diagonal terms of the inverse of $X'X$.

4. Standard variates. Use of correlation matrix. Many of the formulas of section 3 are simplified with the use of some type of standardization. In particular it is possible to reduce the matrix $X'X$ to the matrix R of correlation coefficients by replacing x by t_x/N where $t_x = x/s$. If y is similarly replaced and B by $\mathbf{B}$, then $X'Y = R_{xy}$ and $Y'X = R'_{xy}$, $Y'Y = \Sigma y^2 = 1$ and selected formulas from section 3 become

$$R\mathbf{B} = R_{xy} \tag{18}$$

$$\mathbf{B} = R^{-1}R_{xy} \tag{19}$$

$$r^2_{y.x_1\ldots x_n} = R'_{xy}\mathbf{B} = \mathbf{B}'R\mathbf{B} = R'_{xy}R^{-1}R_{xy}. \tag{20}$$

Classical multiple correlation formulas, determinantal and otherwise, are "covered" by the matrix formulas (20).

5. Matrix presentation of a Doolittle solution. Least squares and correlation constants can also be obtained from the entries of a Doolittle solution. We first outline a matrix description of the Doolittle solution of the equation $AX = G$ with $A = [a_{ij}]$ symmetric and of order n.

Let S_1 be a (n by n) matrix with the first row composed of the elements a_{1j} and all other elements 0. Let T_1 be a similar matrix with first row elements $b_{1j} = a_{1j}/a_{11}$ and all other elements 0. Then $A - S_1'T_1 = A_1 = [a_{ij.1}]$ is a symmetric (n by n) matrix with all elements of the first row and the first column 0.

Next let S_2 be a (n by n) matrix with second row elements $a_{2j.1} = a_{2j} - a_{21}b_{1j}$ and all other elements 0. Let T_2 be a (n by n) matrix with second row elements $b_{2j.1} = a_{2j.1}/a_{22.1}$ and all other elements 0. Then it follows that the matrix

$A_1 - S_2'T_2 = [a_{ij.12}]$ is a symmetric (n by n) matrix with the elements of the first 2 columns and the first 2 rows all 0.

This process is continued through successive steps, an additional row and column being made identically 0 at each step, through n steps. At the end of n steps we have the result.

$$A - S_1'T_1 - S_2'T_2 - \cdots S_n'T_n = 0. \tag{21}$$

This development, when applied to each side of the matrix equation, provides the basis for an equation solving technique which Aitken has called the "method of pivotal condensation" (8) but which the author feels is more adequately characterized as the "method of single division" (9). The Abbreviated Doolittle method can be obtained as an abbreviation of this method. It is not necessary to compute all the elements of the successive matrices $A_1A_2 \cdots$, etc. but only the non-zero elements of the S_1, T_1, S_2, $T_2 \cdots$ etc. matrices.

Consider the so called triangular matrix $S = S_1 + S_2 + S_3 + \cdots + S_n$ with its rows composed of the non-zero rows of the S_j. Consider also the matrix $T = T_1 + T_2 + \cdots + T_n$. Then

$$S'T = S_1'T_1 + S_2'T_2 + \cdots + S_n'T_n \tag{22}$$

$$\text{since } S_i'T_j = 0 \text{ when } i \neq j.$$

It follows that (21) can be written

$$A - S'T = 0. \tag{23}$$

An efficient way of building up these matrices S and T in practice and in making the corresponding transformations on the right side of the equation is the Abbreviated Doolittle method. It is apparent from (23) that the Doolittle method is directed, in part at least, toward the factorization of the symmetric matrix A into two triangular matrices.

It should be noted that these triangular matrices are related by the matrix formula

$$S = DT, \tag{24}$$

where D is the diagonal matrix with diagonal elements a_{11}, $a_{22\cdot1}$, $a_{33\cdot12}$, $\cdots$, $a_{nn\cdot123\cdots n-1}$.

Operations performed on the left of the matrix equations $AX = G$ are also performed on the right side so that the Doolittle technique results in the establishment of the auxiliary matrix equations.

$$SX = SA^{-1}G. \tag{25}$$

$$TX = TA^{-1}G. \tag{26}$$

A simple outline ($n = 3$) of the form of the Abbreviated Doolittle method is presented for the purpose of identifying these matrices. A is symmetric and G is the column matrix $[a_{i4}]$.

$$\begin{array}{ccc|c}
a_{11} & a_{12} & a_{13} & a_{14} \\
- & a_{22} & a_{23} & a_{24} \\
- & - & a_{33} & a_{34} \\
\hline
a_{11} & a_{12} & a_{13} & a_{14} \\
1 & b_{12} & b_{13} & b_{14} \\
\hline
 & a_{22\cdot 1} & a_{23\cdot 1} & a_{24\cdot 1} \\
 & 1 & b_{23\cdot 1} & b_{24\cdot 1} \\
\hline
 & & a_{33\cdot 12} & a_{34\cdot 12} \\
 & & 1 & b_{34\cdot 12}
\end{array}$$

The matrix S is then $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22\cdot 1} & a_{23\cdot 1} \\ 0 & 0 & a_{33\cdot 12} \end{bmatrix}$, the matrix T is

$$\begin{bmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23\cdot 1} \\ 0 & 0 & 1 \end{bmatrix}, \qquad SA^{-1}G \text{ is } \begin{bmatrix} a_{14} \\ a_{24\cdot 1} \\ a_{34\cdot 12} \end{bmatrix}, \qquad TA^{-1}G \text{ is } \begin{bmatrix} b_{14} \\ b_{24\cdot 1} \\ b_{34\cdot 12} \end{bmatrix}.$$

6. Least squares and multiple correlation constants from the Doolittle solution. The inverse of A is needed for many formulas. We set up a technique for solving $AY = I$ simultaneously with $AX = G$. This is indicated symbolically by

$$\begin{array}{c|c|c}
A & G & I \\
\hline
S & SA^{-1}G & SA^{-1} \\
\hline
T & TA^{-1}G & TA^{-1}
\end{array}.$$

It follows at once that

(27) $$(SA^{-1})'(TA^{-1}) = A^{-1}S'TA^{-1} = A^{-1}AA^{-1} = A^{-1}.$$

This matrix multiplication is easily and readily accomplished when the matrices are in the Doolittle form.

Similarly

(28) $(SA^{-1})'(TA^{-1}G) = A^{-1}G = X$, the matrix of solutions of $AX = G$ and

(29) $$(SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G.$$

It is interesting to note further that $(SA^{-1})'$ and T are inverse triangular matrices since

(30) $$(SA^{-1})'T = A^{-1}S'T = I.$$

S' and TA^{-1} have a similar relationship.

In the case of least squares theory $A = X'X$, $G = X'Y$, $X = B$ so that the formulas (27)(28)(29) become

(31) $(SA^{-1})'(TA^{-1}) = A^{-1} = (X'X)^{-1}$.

(32) $(SA^{-1})'(TA^{-1}G) = X = B$.

(33) $(SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G = Y'X(X'X)^{-1}X'Y$

$$= Y'XB = B'X'XB = \underline{Y}\,\underline{Y}.$$

If the normal equations are reduced to standard form $A = R$, $X = \mathbf{B}$, $G = R_{xy}$ and we have

(34) $(SA^{-1})'(TA^{-1}) = A^{-1} = R^{-1}$.

(35) $(SA^{-1})'(TA^{-1}G) = X = \mathbf{B}$.

(36) $(SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G = R'_{xy}R^{-1}R_{xy} = R'_{xy}\mathbf{B} = \mathbf{B}'R\mathbf{B} = r^2_{y \cdot x_1 \cdots x_n}$.

The reader is referred to an earlier paper [3, 457] for an illustration of these techniques.

It should be noted that the solution is a cumulative one in the sense that solutions involving n predicting variables are obtained from solutions involving $n - 1$ predicting variables by the addition of paired products. This is a highly desirable feature as it makes possible direct analyses showing the effect of an added predicting variable.

7. A more general theory—solution of matrix equations by factorization. Examination of the results of section 5 leads one at once to a consideration of a more general theory. The key formula in this development is $A - S'T = 0$ and all subsequent formulas stem from this. Hence if A can be factored into any matrices, S' and T, not necessarily triangular, the results of section 6 follow.

From a practical standpoint it is desired that the factorization process yield, simultaneously, the values S, T, $SA^{-1}G$; $TA^{-1}G$; SA^{-1} and TA^{-1} as the Doolittle method does. But, formally, these can be computed if S and T are known.

8. A "square root" method. A most interesting and practical special case of the above method is that in which the triangular matrices S and T are equal. It appears that a technique based on this property would have some advantages over the Doolittle method since the double rows of the Doolittle solution could be replaced by single rows, while the formulas of sections 5 and 6 are just as applicable. Now such a technique is easily devised. From (23) and (24) we see that

(37) $$A - S'D^{-1}S = 0,$$

where D is a diagonal matrix.

We replace $D^{-\frac{1}{2}}S$ by a new S, $(D^{-\frac{1}{2}}S)'$ by a new S' and have

(38) $$A - S'S = 0.$$

The technique of solution is similar to that of the Doolittle except that the entries $s_{ij\cdots}$ are $a_{ij\cdots}/\sqrt{a_{jj\cdots}}$. These values $s_{ij\cdots}$ are thus geometric means of the values $a_{ij\cdots}$ and $b_{ij\cdots}$.

A simple machine technique is available for computing these entries. In some respects the solution is superior to the Doolittle solution. It is hardly pertinent to the subject matter of this paper to present a detailed discussion of the merits of this method, with the numerical illustrations. This will be done in a later paper.

After arriving at this method by the steps described above, it seemed surprising that such a simple and compact method has not been discovered by some previous worker. Although matrix factorization is not a new subject, I have not found evidence that it has been utilized so directly in the problem of solving matrix equations. The nearest approach I have discovered is the paper by Banachiewicz [10], in which a "square root" method is used in factoring A.

REFERENCES

[1] A. C. AITKEN, *Determinants and Matrices*, Oliver and Boyd, London, 1942.

[2] P. S. DWYER, "The Doolittle technique," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 449–458.

[3] P. S. DWYER, "Recent developments in correlation technique," *Jour. Amer. Stat. Assn.* Vol., 37 (1942), pp. 441–460.

[4] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 1–34.

[5] W. KIRKHAM, "Note on the derivation of the multiple correlation coefficient," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 68–71.

[6] H. M. BACON, "Note on a formula for the multiple correlation coefficient," *Annals of Math. Stat.* Vol. 9, (1938), pp. 227-229.

[7] LOUIS GUTTMAN, "A note on the derivation of a formula for multiple and partial correlation," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 305–308.

[8] A. C. AITKEN "Studies in practical mathematics I. The evaluation, with application, of a certain triple product matrix," *Proc. Royal Soc., Edinburgh*, Vol. 57 (1937), pp. 172-181.

[9] P. S. DWYER, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), pp. 101–129.

[10] T. BANACHIEWICZ, "An outline of the Cracovian algorithm of the method of least squares," *Astr. Jour.* Vol., 50 (1942), pp. 38–41.