BASICS OF MATRIX ALGEBRA WITH STATISTICAL APPLICATIONS
Stevens, 1991

A data matrix is simply a rectangular table of numbers. Matrix algebra is a formal system of manipulating a data matrix that is analogous to ordinary algebra. A number of operations in matrix algebra function in the same way as in ordinary algebra, while other operations are somewhat different. In statistics, where it is common to manipulate data matrices, as the number of variables becomes large, it becomes more and more efficient and decidedly less cumbersome to use matrix algebra rather than ordinary algebra to perform operations. The notation of matrix algebra is also useful, especially in multivariate statistics, since this notation can be used with perfect generality regardless of the number of variables being considered.

One of the most common matrices in statistics is the matrix: $\mathbf{X}_{ij}$, where i = rows (typically of n subjects) and j = columns (typically of different variables). This is a basic Score Matrix. Each entry in the matrix (of which there are i times j entries) is called an element of the matrix. Matrix elements may be zero but may not be empty (blank) for matrix operations to be performed (note the implications this has for missing data treated in matrix form). The following is an example of matrix $\mathbf{X}_{ij}$:

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}$$

In matrix algebra, a single number or a variable which only takes on a single value (a constant) is called a *scalar*. A matrix having only one row or one column is called a *vector*. Therefore, a proper matrix must have at least two rows and two columns of numbers. In matrix algebra, the matrix is typically enclosed in brackets and is denoted by upper-case boldface letters. Lower-case boldface letters refer to vectors. Unless otherwise noted, vectors are assumed to be a column of numbers (rather than a row). The size of a matrix is called its order and refers to the number of rows and columns in the matrix: m X n. The data matrix above is a 3 X 3 matrix, for example. In some matrices, the elements that occur on the diagonal of the matrix have special meaning. It is worth noting then, that the main or principal diagonal of a matrix are the elements going from the upper left corner to the lower right corner. In terms of matrix $\mathbf{X}_{ij}$ above, these are the elements X_{11}, X_{22}, X_{33}.

A matrix with an equal number of rows and columns is called a square matrix. A matrix in which the elements with the same subscripts have the same value, regardless of subscript order, is called a symmetric matrix. In symmetric matrices, corresponding elements on opposite sides of the principal diagonal are equal: $a_{12} = a_{21}$, $a_{35} = a_{53}$, etc.

A diagonal matrix is one in which all elements except those on the principal diagonal are zero, as in the matrix below:

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

MATRIX OPERATIONS

Addition and Subtraction. These operations are performed as in simple algebra but are performed element by element in each matrix:

$$\begin{bmatrix} 2 & 1 \\ 4 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 2+1 & 1+0 \\ 4+2 & 7+3 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 6 & 10 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} + \begin{bmatrix} -2 & -9 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3-(-2) & -1-(-9) \\ -1-1 & 2-4 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ -2 & -2 \end{bmatrix}$$

Only matrices of the size (order) can be added or subtracted. As in simple algebra, addition is commutative (i.e., **A** + **B** = **B** + **A**), associative (i.e., [**A** + **B**] + **C** = **A** + [**B** + **C**]) and distributive: **A** - (**B** + **C**) = **A** - **B** - **C**.

Multiplication. Scalars and matrices may be multiplied. In scalar multiplication, all entries in a matrix are multiplied by the scalar or constant (including scalars that are reciprocals like 1/2, 1/n, etc.):

$$2 \quad X \quad \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 2 & -2 \end{bmatrix}$$

In order to multiply two matrices, the column width of the matrix on the left must be the same size as the row size of the matrix on the right:

Size must match

4 X 2 times 2 X 6

Size of product matrix

Thus, to be multiplied, two matrices must match in terms of the "interior" dimensions of the matrices. The resulting product matrix will be equal in size to the "exterior" dimensions of the original matrices. In the example above, the product of the 4 X 2 matrix times the 2 X 6 matrix will be a 4 X 6 matrix.

Matrix multiplication is accomplished by moving across the rows of the first matrix while moving down the columns of the second matrix. Thus the first element of the product matrix is formed by taking element X_{11} multiplied by element Y_{11} and summed with the product of X_{12} and Y_{21}. This process is illustrated below:

$$\underset{\mathbf{X}}{\begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}} \underset{\mathbf{Y}}{\begin{bmatrix} 1 & 3 & -3 \\ 2 & 0 & 2 \end{bmatrix}} = \begin{bmatrix} (2)(1)+(3)(2) & (2)(3)+(3)(0) & (2)(-3)+(3)(2) \\ (1)(1)+(-1)(2) & (1)(3)+(-1)(0) & (1)(-3)+(-1)(2) \end{bmatrix}$$

Another way to visualize this process (which seems very unusual compared to ordinary algebra), is to think of it as a multiplication that occurs after turning the first matrix 90° clockwise:

$$\begin{bmatrix} 1 & 3 & -3 \\ 2 & 0 & 2 \end{bmatrix}$$

Each column in the first matrix is then multiplied times the element in the same row of the first column of the second matrix. This is repeated column by column.

In matrix algebra pre- and post-multiplication is not the same thing at all:

$$\begin{bmatrix} 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 12$$

1 X 3 times 3 X 1 = 1 X 1

$$\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \begin{bmatrix} 4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 2 & 2 \\ 12 & 3 & 3 \\ 4 & 1 & 1 \end{bmatrix}$$

3 x 1 times 1 X 3 = 3 X 3

Therefore, matrix multiplication is not commutative: **AB** ≠ **BA**, and it is very important to distinguish pre- from post-multiplication. However, as in ordinary algebra, matrix multiplication is associative: **ABC** = **A**(**BC**) = (**AB**)**C** nad is distributive: **Q**(**M** + **Z**) = **QM** + **QZ**.

Identity Matrices. A matrix which has no effect on another matrix when the two are multiplied is called an identity matrix, **I**. For example:

$$\underset{\mathbf{A}}{\begin{bmatrix} 0 & 2 \\ -1 & 4 \end{bmatrix}} \underset{\mathbf{I}}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = \underset{\mathbf{A}}{\begin{bmatrix} 0 & 2 \\ -1 & 4 \end{bmatrix}}$$

There is a matrix **I** such that for any matrix **A**, the following is true: **AI** = **IA** = **A**.

Transposing Matrices. The transpose of a matrix is a rotation of the matrix in which every row of the original matrix becomes a column of the transposed matrix. The numbers remain the same. Every matrix has a transpose denoted by a prime (') symbol. The transpose of **A** is **A'**:

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 4 \end{bmatrix} \qquad \mathbf{A'} = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Transposing a transposed matrix returns the matrix to its original form, (**A'**)' = **A**. The most common use of the transpose is to allow multiplication of matrices when interior dimensions do not match. For example, take the score matrix, **X**, which is n X p, composed of n subjects and p variables:

$$\mathbf{X} = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

If we wanted to multiply the 4 x 3 matrix by itself, the interior dimensions would not match and multiplication would not be possible. But if the original matrix is transposed and then premultiplied, the operation becomes possible:

$$\underset{\substack{\mathbf{X'} \\ 3 \text{ X } 4}}{\begin{bmatrix} 3 & 1 & 1 & 3 \\ 1 & 1 & 2 & 2 \\ 1 & 0 & 2 & 1 \end{bmatrix}} \underset{\substack{\mathbf{X} \\ 4 \text{ X } 3}}{\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}} = \underset{\substack{\mathbf{X'X} \\ 3 \text{ X } 3}}{\begin{bmatrix} 20 & 12 & 8 \\ 12 & 10 & 7 \\ 8 & 7 & 6 \end{bmatrix}}$$

This operation is one of the most common in statistics, and the **X'X** matrix is a particularly useful one. Note what happens in the matrix multiplication process: the score of subject 1 on variable 1 (element X_{11} in **X'**) is multiplied times itself (X_{11} in **X**), the product is then added to the square of the second subject's score on variable 1, and so on through the first row and first column. In the product matrix, **X'X**, this yields a first element that is the sums of squares for the first variable. In forming the second element of the product matrix, scores on the first variable are multiplied times scores on the second variable to form crossproducts. Thus, the premultiplication of the score matrix, **X**, by its transpose, **X'**, results in a matrix, **X'X**, which is called a sums of squares, crossproducts matrix or SSCP matrix. Symbolically:

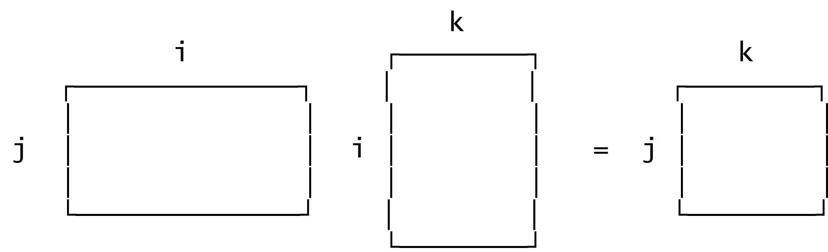
$$\underset{\mathbf{X'}}{\begin{bmatrix} X_{11} & X_{21} & X_{31} & X_{41} \\ X_{12} & X_{22} & X_{32} & X_{42} \\ X_{13} & X_{23} & X_{33} & X_{43} \end{bmatrix}} \underset{\mathbf{X}}{\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \end{bmatrix}} = \underset{\mathbf{X'X}}{\begin{bmatrix} \Sigma X^2_1 & \Sigma X_1X_3 & \Sigma X_1X_3 \\ \Sigma X_2X_1 & \Sigma X^2_2 & \Sigma X_2X_3 \\ \Sigma X_3X_1 & \Sigma X_3X_2 & \Sigma X^2_3 \end{bmatrix}}$$

Statistical Formulae in Matrix Notation

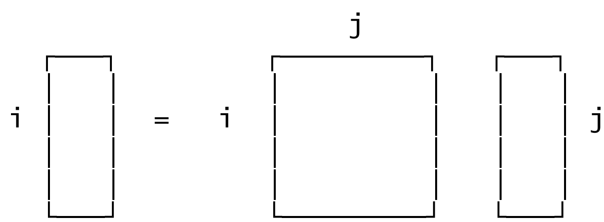
A wide variety of statistical problems can be represented and solved using one basic general form of matrix multiplication: **C** = **AB**. This matrix algebra formulation is equivalent to the following formula in ordinary algebra:

$$C_{jk} = \Sigma \; A_{ji} \; B_{ik} \qquad (1)$$

And we can give a general representation of the matrices that would be involved using boxes as below. Compare the subscripts of A and B in formula 1 to the dimensions of the matrices. Now look at the subscripts for C in formula 1. Note that the "i" dimension in the original matrices has "collapsed", leaving only "jk" in the product matrix. This reveals something that you probably observed already in the product matrix, **X'X**, above. That is, in matrix multiplication, the interior dimensions of the original matrices are summed and do not appear in the product matrix.



For example, take the following instance of formula 1: $Y_i = \Sigma \; x_{ij} \; b_j$
Note that b has only one subscript indicating that b will be a vector rather than a matrix. The matrix representation of the same formula is: **y** = **Xb**, and the computations can be represented as:



Taking a particular example with i = 3 and j = 2 we have:

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} x_{11}b_1 + x_{12}b_2 \\ x_{21}b_1 + x_{22}b_2 \\ x_{31}b_1 + x_{32}b_2 \end{bmatrix} = \begin{bmatrix} \Sigma x_1 b \\ \Sigma x_2 b \\ \Sigma x_3 b \end{bmatrix}$$

You may recognize this procedure as the process of taking the sums of squares of deviations for two predictors, X_1 and X_2, and multiplying by beta weights to obtain predicted scores on the criterion, Y.

<u>Obtaining Simple Sums</u>. In matrix algebra, simple sums of the form ΣX_{ij} can be obtained through multiplication of the data matrix times a vector consisting

of 1's for each element in the vector. This vector is denoted as **1**. Take as an example a matrix consisting of i subjects, each of which has a score on each of j items of a test. This results in an i X j data matrix, **X**. To find the sum across the j items for each subject, postmultiply **X** by a vector of ones, **1**: $\mathbf{x}_i = \mathbf{X}_{ij}\mathbf{1}_j$:

$$\begin{bmatrix} 3 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$$

If summation across rows of the data matrix is desired, the vector, **1**, is transposed and premultiplied: $\mathbf{x}_j = (\mathbf{1'}_i)(\mathbf{X}_{ij})$. To sum over the whole matrix, pre- and postmultiply. This results in the grand sum of all the scores (a scalar): $\mathbf{x} = (\mathbf{1'}_i)(\mathbf{X}_{ij})(\mathbf{1}_j)$:

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 12$$

<u>Multiplying and Dividing by Constants</u>. To multiply all elements in a matrix by a single value, on uses scalar multiplication as described earlier: k**X** or **X**k. Division can be performed by multiplying by the reciprocal of the constant: (1/k)(**X**), for example. However, the latter is usually represented as k^{-1}.

It is often useful, however, to multiply or divide particular rows or columns of the data matrix by different constants. This can be accomplished using a diagonal matrix. For example, each column of a data matrix can multiplied by a different constant by postmultiplying a diagonal matrix containing a constant for each column of the data matrix:

A common instance of this procedure is the transformation of variables. If **X** is composed of deviation scores, (X-X), and each d_j is the standard deviation of the respective X_j, then the multiplication above would transform the deviation scores into z scores.

The same procedure can be accomplished with respect to rows of a data matrix by premultiplying by the diagonal matrix: **T** = **DX**.

<u>Computing Means</u>. Given a score matrix, $\mathbf{X}_{ij}$, containing the scores of i = 1 to n subjects on j variables, the means of each variable are computed as:

$$X_j = \Sigma X_{ij} / n$$

Given that the summation in this expression is across the rows of the i X j score matrix, the same operations can be performed in matrix algebra by premultiplying by **1'**, and postmultiplying by the constant 1/n:

$$\mathbf{x'}_j = \mathbf{1'Xn}^{-1}$$

For example:

$$\begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{vmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{vmatrix} \begin{vmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \end{vmatrix}$$

$$\begin{bmatrix} \cdots & & \\ X_{31} & X_{32} & X_{33} \end{bmatrix} \begin{bmatrix} \cdots & & \\ 0 & 0 & 1/3 \end{bmatrix}$$

$$\mathbf{x'} = \mathbf{1'} \quad \mathbf{X} \quad \mathbf{n}^{-1}$$

$$\begin{bmatrix} 4.67 & 3.33 & 2.67 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 8 & 6 & 5 \end{bmatrix} \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

Computing Deviation Scores. When a column vector is postmultiplied by a row vector an *outer product* of the two vectors is formed. This principle can be used to construct a matrix, $\mathbf{X}_j$, in which each column j is a constant X_j: $\mathbf{X}_j = \mathbf{1x}_j$:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 4.67 & 3.33 & 2.67 \end{bmatrix} = \begin{bmatrix} 4.67 & 3.33 & 2.67 \\ 4.67 & 3.33 & 2.67 \\ 4.67 & 3.33 & 2.67 \end{bmatrix}$$

$$\mathbf{1} \quad \mathbf{x} = \mathbf{X}$$

This product matrix can then be used to transform a score matrix into deviation score form. If D is a matrix of deviation scores, then $D_{ij} = X_{ij} - X_j$, or in matrix form: $\mathbf{D}_{ij} = \mathbf{X}_{ij} - \mathbf{X}_j$. (If we substitute the matrix formula for calculating means above, then deviation scores can be calulated from the original score matrix, $\mathbf{X}$, by: $\mathbf{D}_{ij} = \mathbf{X}_{ij} - \mathbf{11'Xn}^{-1}$). Using $\mathbf{X}_j$ we calculate $\mathbf{D}_{ij}$ as:

$$\begin{bmatrix} -2.67 & -0.33 & -1.67 \\ -0.67 & -2.33 & -0.67 \\ 3.33 & 2.67 & 2.33 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 8 & 6 & 5 \end{bmatrix} - \begin{bmatrix} 4.67 & 3.33 & 2.67 \\ 4.67 & 3.33 & 2.67 \\ 4.67 & 3.33 & 2.67 \end{bmatrix}$$

$$\mathbf{D} = \mathbf{X} - \mathbf{X}$$

Now we can calculate another matrix that is useful in statistics by using a diagonal matrix, $\mathbf{S}^{-1}$. If this diagonal matrix is composed of the reciprocals of the standard deviations of each of the j variables, then by postmultiplying $\mathbf{D}_{ij}$ by $\mathbf{S}^{-1}$, we can divide each deviation score by its standard deviation. Of course, dividing a deviation from the mean by the standard deviation results in a standard or z-score. Therefore, to obtain z-scores: $\mathbf{Z}_{ij} = (\mathbf{D}_{ij})(\mathbf{S}^{-1})$:

$$\begin{bmatrix} -1.07 & -0.16 & -0.98 \\ -0.27 & -1.13 & -0.39 \\ +1.34 & +1.30 & +1.37 \end{bmatrix} = \begin{bmatrix} -2.67 & -0.33 & -1.67 \\ -0.67 & -2.33 & -0.67 \\ 3.33 & 2.67 & 2.33 \end{bmatrix} \begin{bmatrix} 1/2.49 & 0 & 0 \\ 0 & 1/2.05 & 0 \\ 0 & 0 & 1/1.70 \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{D} \quad \mathbf{S}^{-1}$$

Variance-Covariance Matrices. The matrix of deviation scores, $\mathbf{D}_{ij}$, can also be used to calculate the variances and covariances of each of the variables. In ordinary algebra: $VAR_j = \Sigma x^2_{ij} / n-1$, and $COV_{jk} = \Sigma x_{ij} x_{ik} / n-1$. To accomplish these operations in matrix algebra, the deviation score matrix, $\mathbf{D}_{ij}$, is premultiplied by its transpose, $\mathbf{D'}_{ij}$. The resulting product matrix is a sums of squares-crossproducts matrix (SSCP) and will contain the numerator of variances down the diagonal and the numerator of the covariances elsewhere. To obtain the variance-covariance matrix it is then necessary to divide the SSCP matrix by n-1, or in matrix terms, postmultiply by $(\mathbf{n-1})^{-1}$. This gives us the following matrix formulation for calculating a variance-covariance natrix from a matrix of deviation scores: $\mathbf{V} = (\mathbf{D'D})(\mathbf{n-1})^{-1}$ For example:

$$\underset{\mathbf{D'}}{\begin{bmatrix} -2.67 & -0.67 & 3.33 \\ -0.33 & -2.33 & 2.67 \\ -1.67 & -0.67 & 2.33 \end{bmatrix}} \underset{\mathbf{D}}{\begin{bmatrix} -2.67 & -0.33 & -1.67 \\ -0.67 & -2.33 & -0.67 \\ 3.33 & 2.67 & 2.33 \end{bmatrix}} = \underset{\mathbf{D'D}}{\begin{bmatrix} 18.67 & 11.33 & 12.67 \\ 11.33 & 12.67 & 8.33 \\ 12.67 & 8.33 & 8.67 \end{bmatrix}}$$

$$\underset{\mathbf{D'D}}{\begin{bmatrix} 18.67 & 11.33 & 12.67 \\ 11.33 & 12.67 & 8.33 \\ 12.67 & 8.33 & 8.67 \end{bmatrix}} \underset{(\mathbf{n-1})^{-1}}{\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}} = \underset{\mathbf{V}}{\begin{bmatrix} 9.34 & 5.67 & 6.34 \\ 5.67 & 6.34 & 4.17 \\ 6.34 & 4.17 & 4.34 \end{bmatrix}}$$

So, the variances of variables one through three are 9.34, 6.34, and 4.34, respectively. The covariance of varaibles 2 and 1 is 5.67, etc. The variance-covariance matrix, **V**, is a workhorse in statistics and forms the basis for the calculation of many common procedures.

<u>Practice</u>. Matrix algebra is not nearly as imposing as it appears at first glance. A little practice with some of the operations described above will build your confidence. Try the following exercises:

1. Can a 3 X 4 matrix be multiplied by another 3 X 4 matrix? How?

2. Write a matrix expression that will perform a summation over k in a j X k matrix.

3. How would you obtain the grand sum of the matrix in question 1?

4. Use the numbers in matrix **V** above as the elements of an i X j score matrix. Compute the means of the j variables using the matrix formulation.

5. After computing means in question 3, see if you can carry the matrix process all the way through to computation of a variance-covariance matrix.