Matrix Multiplication

Matrix multiplication is an operation with properties quite different from its scalar counterpart.

To begin with, *order matters* in matrix multiplication. That is, the matrix product **AB** need not be the same as the matrix product **BA**. Indeed, the matrix product **AB** might be well-defined, while the product **BA** might not exist.

Definition (Conformability for Matrix Multiplication).

${}_p\mathbf{A}_q$ *and* ${}_r\mathbf{B}_s$ *are conformable for matrix multiplication as* $\mathbf{AB}$ *if and only if* $q = r$.

Definition (Matrix Multiplication). *Let* ${}_p\mathbf{A}_q = \left\{a_{ij}\right\}$ *and* ${}_q\mathbf{B}_s = \left\{b_{ij}\right\}$.

Then ${}_p\mathbf{C}_s = \mathbf{AB} = \left\{c_{ik}\right\}$ *where*

$$c_{ik} = \sum_{j=1}^{q} a_{ij} b_{jk} \qquad (1)$$

Example (The Row by Column Method). The meaning of the formal definition of matrix multiplication might not be obvious at first glance. Indeed, there are several ways of thinking about matrix multiplication.

The first way, which I call the "*row by column approach*," works as follows. Visualize ${}_p\mathbf{A}_q$ as a set of p *row vectors* and ${}_q\mathbf{B}_s$ as a set of s *column vectors*. Then if $\mathbf{C} = \mathbf{AB}$, element c_{ik} of $\mathbf{C}$ is the scalar product (i.e., the sum of cross products) of the ith row of $\mathbf{A}$ with the kth column of $\mathbf{B}$.

For example, let $\mathbf{A} = \left[\begin{array}{ccc} 2 & 4 & 6 \\ \hline 5 & 7 & 1 \\ \hline 2 & 3 & 5 \end{array}\right]$, and let

$$\mathbf{B} = \left[\begin{array}{c|c} 4 & 1 \\ 0 & 2 \\ 5 & 1 \end{array}\right]$$

Then $\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 38 & 16 \\ 25 & 20 \\ 33 & 13 \end{bmatrix}$.

The following are some key properties of matrix multiplication:

1) Associativity.

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \qquad (2)$$

2) Not generally commutative. That is, often $\mathbf{AB} \neq \mathbf{BA}$.

3) Distributive over addition and subtraction.

$$\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB} \qquad (3)$$

4) Assuming it is conformable, the identity matrix **I** functions like the number 1, that is

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \qquad (4)$$

5) $\mathbf{AB} = \mathbf{0}$ does not necessarily imply that either $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

Several of the above results are surprising, and result in negative transfer for beginning students as they attempt to reduce matrix algebra expressions.

Example (A Null Matrix Product). The following example shows that one can, indeed, obtain a null matrix as the product of two non-null matrices. Let

$$\mathbf{a}' = \begin{bmatrix} 6 & 2 & 2 \end{bmatrix}, \text{ and let } \mathbf{B} = \begin{bmatrix} -8 & 12 & 12 \\ 12 & -40 & 4 \\ 12 & 4 & -40 \end{bmatrix}.$$

Then $\mathbf{a}'\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$.

Definition (Pre-multiplication and Post-multiplication).

When we talk about the “product of matrices **A** and **B**,” it is important to remember that **AB** and **BA** are usually not the same. Consequently, it is common to use the terms “pre-multiplication” and “post-multiplication.” When we say “**A** is post-multiplied by **B**,” or “**B** is pre-multiplied by **A**,” we are referring to the product **AB** . When we say “**B** is post-multiplied by **A**,” or “**A** is pre-multiplied by **B**,” we are referring to the product **BA** .

Matrix Transposition

“Transposing” a matrix is an operation which plays a very important role in multivariate statistical theory. The operation, in essence, switches the rows and columns of a matrix.

Definition (Matrix Transposition).

Let ${}_p\mathbf{A}_q = \{a_{ij}\}$ *. Then the transpose of* $\mathbf{A}$*, denoted* $\mathbf{A}'$ *or* $\mathbf{A}^T$*, is defined as*

$$\mathbf{B} = {}_q\mathbf{A}'_p = \{b_{ij}\} = \{a_{ji}\} \qquad (5)$$

Example (Matrix Transposition).

$$Let\ \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix}.\ Then\ \mathbf{A}' = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 5 \end{bmatrix}$$

Properties of Matrix Transposition.

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

$$(c\mathbf{A})' = c\mathbf{A}'$$

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

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

A square matrix **A** is symmetric if and only if

$$\mathbf{A} = \mathbf{A}'$$

Partitioning of Matrices

In many theoretical discussions of matrices, it will be useful to conceive of a matrix as being composed of sub-matrices. When we do this, we will “partition” the matrix symbolically by breaking it down into its components. The components can be either matrices or scalars.

Example. *In simple multiple regression, where there is one criterion variable y and p predictor variables in the vector* **x**, *it is common to refer to the correlation matrix of the entire set of variables using partitioned notation. So we can write*

$$\mathbf{R} = \begin{bmatrix} 1 & \mathbf{r}'_{y\mathbf{x}} \\ \mathbf{r}_{\mathbf{x}y} & \mathbf{R}_{\mathbf{xx}} \end{bmatrix} \qquad (6)$$

Order of a Partitioned Form

We will refer to the "order" of the "partitioned form" as the number of rows and columns in the partitioning, which is distinct from the number of rows and columns in the matrix being represented. For example, suppose there were $p = 5$ predictor variables in the example of Equation (6). Then the matrix $\mathbf{R}$ is a 6×6 matrix, but the example shows a " 2×2 partitioned form."

When matrices are partitioned *properly*, it is understood that "pieces" that appear to the left or right of other pieces have the same number of rows, and pieces that appear above or below other pieces have the same number of columns. So, in the above example, $\mathbf{R_{xx}}$, appearing to the right of the $p \times 1$ column vector $\mathbf{r}_{\mathbf{x}y}$, must have p rows, and since it appears below the $1 \times p$ row vector $\mathbf{r}'_{y\mathbf{x}}$, it must have p columns. Hence, it must be a $p \times p$ matrix.

Linear Combinations of Matrix Rows and Columns

We have already discussed the “row by column” conceptualization of matrix multiplication. However, there are some other ways of conceptualizing matrix multiplication that are particularly useful in the field of multivariate statistics. To begin with, we need to enhance our understanding of the way matrix multiplication and transposition works with partitioned matrices.

Definition. (Multiplication and Transposition of Partitioned Matrices).

1. To transpose a partitioned matrix, treat the sub-matrices in the partition as though they were elements of a matrix, but transpose each sub-matrix. The transpose of a $p \times q$ partitioned form will be a $q \times p$ partitioned form.

2. To multiply partitioned matrices, treat the sub-matrices as though they were elements of a matrix. The product of $p \times q$ and $q \times r$ partitioned forms is a $p \times r$ partitioned form.

Some examples will illustrate the above definition.

Example (Transposing a Partitioned Matrix).

Suppose $\mathbf{A}$ is partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}. \text{ Then } \mathbf{A}' = \begin{bmatrix} \mathbf{C}' & \mathbf{E}' & \mathbf{G}' \\ \mathbf{D}' & \mathbf{F}' & \mathbf{H}' \end{bmatrix}$$

Example (Product of Two Partitioned Matrices).

Suppose $\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix}$.

Then (assuming conformability)

$$\mathbf{AB} = \mathbf{XG} + \mathbf{YH}$$

Example (Linearly Combining Columns of a Matrix).

Consider an $N \times p$ matrix $\mathbf{X}$, containing the scores of N persons on p variables. One can conceptualize the matrix as a set of p column vectors. In "partitioned matrix form," we can represent $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_\mathbf{2} & \mathbf{x}_\mathbf{3} & \cdots & \mathbf{x}_\mathbf{p} \end{bmatrix}$$

Now suppose one were to post-multiply $\mathbf{X}$ with a $p \times 1$ vector $\mathbf{b}$. The product is a $N \times 1$ column vector:

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

$$= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_p \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_p \end{bmatrix}$$

$$= b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3 + \cdots + b_p\mathbf{x}_p$$

Example (Computing Difference Scores).

Suppose the matrix **X** consists of a set of scores on two variables, and you wish to compute the difference scores on the variables.

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

$$= \begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} +1 \\ -1 \end{bmatrix} = \begin{bmatrix} 10 \\ -2 \\ 0 \end{bmatrix}$$

Example. (Computing Course Grades).

$$\begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 73\frac{1}{3} \\ 78\frac{1}{3} \\ 64 \end{bmatrix}$$

Example. (Linearly Combining Rows of a Matrix).

Suppose we view the $p \times q$ matrix $\mathbf{X}$ as being composed of p row vectors. If we pre-multiply $\mathbf{X}$ with a $1 \times p$ row vector $\mathbf{b}'$, the elements of $\mathbf{b}'$ are linear weights applied to the rows of $\mathbf{X}$.

Sets of Linear Combinations

There is, of course, no need to restrict oneself to a single linear combination of the rows and columns of a matrix. To create more than one linear combination, simply add columns (or rows) to the post-multiplying (or pre-multiplying) matrix!

$$\begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 150 & 10 \\ 156 & -2 \\ 128 & 0 \end{bmatrix}$$

Example (Extracting a Column from a Matrix).

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

Definition (Selection Vector). The selection vector $\mathbf{s}_{[i]}$ is a vector with all elements zero except the ith element, which is 1. To extract the ith column of $\mathbf{X}$, post-multiply by $\mathbf{s}_{[i]}$, and to extract the ith row of $\mathbf{X}$, pre-multiply by $\mathbf{s}'_{[i]}$.

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 5 \end{bmatrix}$$

Example (Exchanging Columns of a Matrix).

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 5 & 2 \\ 6 & 3 \end{bmatrix}$$

Example (Rescaling Rows or Columns).

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 12 \\ 4 & 15 \\ 6 & 18 \end{bmatrix}$$

Example (Using Two Selection Vectors to Extract a Matrix Element).

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4$$

Matrix Algebra of Some Sample Statistics

Converting to Deviation Scores

Suppose **x** is an $N \times 1$ vector of scores for N people on a single variable. We wish to transform the scores in to *deviation score form*. (In general, we will find this a source of considerable convenience.) To accomplish the deviation score transformation, the arithmetic mean $\bar{X}_{\bullet}$, must be subtracted from each score in **x**.

Let $\mathbf{1}$ be a $N \times 1$ vector of ones.
Then

$$\sum_{i=1}^{N} X_i = \mathbf{1}'\mathbf{x}$$

and

$$\bar{X}_{\bullet} = (1/N)\sum_{i=1}^{N} X_i = (1/N)\mathbf{1}'\mathbf{x}$$

To transform to deviation score form, we need to subtract $\bar{X}_{\bullet}$ from every element of **x**. We need

$$
\begin{aligned}
\mathbf{x}^{*} &= \mathbf{x} - \mathbf{1}(\bar{X}_{\bullet}) \\
&= \mathbf{x} - \mathbf{1}\mathbf{1}'\mathbf{x}/N \\
&= \mathbf{x} - (\mathbf{1}\mathbf{1}'/N)\mathbf{x} \\
&= \mathbf{I}\mathbf{x} - (\mathbf{1}\mathbf{1}'/N)\mathbf{x} \\
&= \mathbf{I}\mathbf{x} - \mathbf{P}\mathbf{x} \\
&= (\mathbf{I} - \mathbf{P})\mathbf{x} \\
&= \mathbf{Q}\mathbf{x}
\end{aligned}
$$

Example

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

Note that the ith row of $\mathbf{Q}$ gives you a linear combination of the N scores for computing the ith deviation score.

Properties of the Q Operator

Definition (Idempotent Matrix).

A matrix $\mathbf{C}$ is idempotent if $\mathbf{CC} = \mathbf{C}^{\mathbf{2}} = \mathbf{C}$

Theorem. If $\mathbf{C}$ is idempotent and $\mathbf{I}$ is a conformable identity matrix, then $\mathbf{I} - \mathbf{C}$ is also idempotent.

Proof. To prove the result, we need merely show that $(\mathbf{I} - \mathbf{C})^2 = (\mathbf{I} - \mathbf{C})$. This is straightforward.

Properties of the Q Operator

$$\begin{aligned}
(\mathbf{I}-\mathbf{C})^2 &= (\mathbf{I}-\mathbf{C})(\mathbf{I}-\mathbf{C}) \\
&= \mathbf{I}^2 - \mathbf{C}\mathbf{I} - \mathbf{I}\mathbf{C} + \mathbf{C}^2 \\
&= \mathbf{I} - \mathbf{C} - \mathbf{C} + \mathbf{C} \\
&= \mathbf{I} - \mathbf{C}
\end{aligned}$$

Properties of the Q Operator

Class Exercise. Prove that if a matrix $\mathbf{A}$ is symmetric, so is $\mathbf{AA}'$.

Class Exercise. From the preceding, prove that if a matrix $\mathbf{A}$ is symmetric, then for any scalar c, the matrix $c\mathbf{A}$ is symmetric.

Class Exercise. If matrices $\mathbf{A}$ and $\mathbf{B}$ are both symmetric and of the same order, then $\mathbf{A}+\mathbf{B}$ and $\mathbf{A}-\mathbf{B}$ must be symmetric.

Properties of the Q Operator

Recall that $\mathbf{P} = \mathbf{11}'/N$ is an $N \times N$ symmetric matrix with each element equal to $1/N$. $\mathbf{P}$ is also idempotent. (See handout.)

It then follows that $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ is also symmetric and idempotent. (Why? C.P.)

The Sample Variance

If $\mathbf{x}^*$ has scores in deviation score form, then

$$S_X^2 = 1/(N-1)\mathbf{x}^{*\prime}\mathbf{x}^*$$

The Sample Variance

If scores in **x** are not in deviation score form, we may use the **Q** operator to convert it into deviation score form first. Hence, in general,

$$
\begin{aligned}
S_X^2 &= 1/(N-1)\mathbf{x}'\mathbf{Q}'\mathbf{Q}\mathbf{x} \\
&= 1/(N-1)\mathbf{x}'\mathbf{Q}\mathbf{Q}\mathbf{x} \\
&= 1/(N-1)\mathbf{x}'\mathbf{Q}\mathbf{x}
\end{aligned}
$$

The Sample Covariance

Do you understand each step below? Remember that $\mathbf{Q} = \mathbf{Q}' = \mathbf{QQ} = \mathbf{Q}'\mathbf{Q} = \mathbf{QQ}'$

$$
\begin{aligned}
S_{XY} &= 1/(N-1)\mathbf{x}'\mathbf{Qy} \\
&= 1/(N-1)\mathbf{x}'\mathbf{y}^{*} \\
&= 1/(N-1)\mathbf{x}'\mathbf{Q}'\mathbf{y} \\
&= 1/(N-1)\mathbf{x}^{*\prime}\mathbf{y} \\
&= 1/(N-1)\mathbf{x}'\mathbf{Q}'\mathbf{Qy} \\
&= 1/(N-1)\mathbf{x}^{*\prime}\mathbf{y}^{*}
\end{aligned}
$$

Notational Conventions

In what follows, we will generally assume, unless explicitly stated otherwise, that our data matrices have been transformed to deviation score form. (The operator discussed above will accomplish this simultaneously for the case of scores of N subjects on several, say p, variates.) For example, consider a data matrix ${}_N\mathbf{X}_p$, whose p columns are the scores of N subjects on p different variables. If the columns of $\mathbf{X}$ are in raw score form, the matrix $\mathbf{Qx}$ will have p columns of deviation scores. Why?

Notational Conventions

We shall concentrate on results in the case where is in "column variate form," i.e., is an $N \times p$ matrix. Equivalent results may be developed for "row variate form" $p \times N$ data matrices which have the N scores on p variables arranged in p rows. The choice of whether to use row or column variate representations is arbitrary, and varies in books and articles.

The Variance-Covariance Matrix

$$\mathbf{S}_{\mathbf{XX}} = 1/(N-1)\mathbf{X}'\mathbf{Q}\mathbf{X}$$

If we assume **X** is in deviation score form, then

$$\mathbf{S}_{\mathbf{XX}} = 1/(N-1)\mathbf{X}'\mathbf{X}$$

(Note: Some authors call $\mathbf{S}_{\mathbf{XX}}$ a “covariance matrix.”) (Why would they do this?)

Diagonal Matrices

Diagonal matrices have special properties, and we have some special notations associated with them. We use the notation diag(**X**) to signify a diagonal matrix with diagonal entries equal to the diagonal elements of **X**.

We use “power notation” with diagonal matrices, in the following sense: Let **D** be a diagonal matrix. Then $\mathbf{D}^c$ is a diagonal matrix composed of the entries of **D** raised to the *c* power.

Correlation Matrix

For p variables in the data matrix $\mathbf{X}$, the *correlation matrix* $\mathbf{R_{XX}}$ is a $p \times p$ symmetric matrix with typical element r_{ij} equal to the correlation between variables i and j. Of course, the diagonal elements of this matrix represent the correlation of a variable with itself, and are all equal to 1.

Correlation Matrix

$$\mathbf{R}_{\mathbf{XX}} = \mathbf{D}^{-1/2}\mathbf{S}_{\mathbf{XX}}\mathbf{D}^{-1/2}$$

(Cross-) Covariance Matrix

Assume **X** and **Y** are in deviation score form. Then

$$\mathbf{S}_{\mathbf{XY}} = 1/(N-1)\mathbf{X}'\mathbf{Y}$$

Variance-Covariance of Linear Combinations

Theorem (Linear Combinations of Deviation Scores). Given $\mathbf{X}$, a data matrix in column variate deviation score form. Any linear composite $\mathbf{Y} = \mathbf{X}\mathbf{b}$ will also be in deviation score form.

Variance and Covariance of Linear Combinations

Theorem. (Variance-Covariance of Linear Combinations).

a) If $\mathbf{X}$ has variance-covariance matrix $\mathbf{S}_{\mathbf{xx}}$, then the linear combination $\mathbf{y} = \mathbf{X}\mathbf{b}$ has variance $\mathbf{b}'\mathbf{S}_{\mathbf{XX}}\mathbf{b}$.

b) The *set of linear combinations* $\mathbf{Y} = \mathbf{X}\mathbf{B}$ has variance-covariance matrix $\mathbf{S}_{\mathbf{YY}} = \mathbf{B}'\mathbf{S}_{\mathbf{XX}}\mathbf{B}$.

c) Two sets of linear combinations $\mathbf{W} = \mathbf{XB}$ and $\mathbf{M} = \mathbf{YC}$ have covariance matrix $\mathbf{S}_{\mathbf{WM}} = \mathbf{B}'\mathbf{S}_{\mathbf{XY}}\mathbf{C}$.

Trace of a Square Matrix

Definition (Trace of a Square Matrix).

The trace of a $p \times p$ square matrix $\mathbf{A}$ is

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{p} a_{ii}$$

Properties of the Trace

1. $\mathrm{Tr}(\mathbf{A}+\mathbf{B}) = \mathrm{Tr}(\mathbf{A}) + \mathrm{Tr}(\mathbf{B})$
2. $\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{A}')$
3. $\mathrm{Tr}(c\mathbf{A}) = c\,\mathrm{Tr}(\mathbf{A})$
4. $\mathrm{Tr}(\mathbf{A}'\mathbf{B}) = \sum_{i}\sum_{j} a_{ij}b_{ij}$
5. $\mathrm{Tr}(\mathbf{E}'\mathbf{E}) = \sum_{i}\sum_{j} e_{ij}^2$
6. The "cyclic permutation rule"

$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA})$$