Econ61001 Econometric Methods

Lecture 2 Matrices - 1

Len Gill

Semester 1 2013 - 2014

Contents

1 Matrices

In the PreSession Maths course, a matrix was defined as follows:

A matrix is a rectangular array of numbers enclosed in parentheses, conventionally denoted by a capital letter. The number of rows (say m) and the number of columns (say n) determine the order of the matrix ($m \times n$). Two examples were given:

$$P = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 1 & 5 \end{bmatrix}, \quad Q = \begin{bmatrix} 2 & 3 \\ 4 & 3 \\ 1 & 5 \end{bmatrix},$$

matrices of dimensions 2×3 and 3×2 respectively.

Why study matrices for econometrics? Basically because a data set of several variables, e.g. on the weights and heights of 12 students, can be thought of as a matrix:

$$D = \begin{bmatrix} 155 & 70 \\ 150 & 63 \\ 180 & 72 \\ 135 & 60 \\ 156 & 66 \\ 168 & 70 \\ 178 & 74 \\ 160 & 65 \\ 132 & 62 \\ 145 & 67 \\ 139 & 65 \\ 152 & 68 \end{bmatrix}$$

The properties of matrices can then be used to facilitate answering all the usual questions of econometrics - list not given here!

Calculations with matrices with explicit numerical elements, as in the examples above is called matrix *arithmetic*. Matrix *algebra* is the algebra of matrices where the elements are not made explicit: this is what is really required for econometrics, as we shall see.

As an example of this, a 2×3 matrix might be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix},$$

and would equal P above if the collection of a_{ij} were given appropriate numerical values.

A general $m \times n$ matrix A can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

There is also a *typical element* notation for matrices:

$$A = \left\| a_{ij} \right\|, \quad i = 1, ..., m, j = 1, ..., n,$$

so that a_{ij} is the element at the intersection of the ith row and jth column in A.

When $m \neq n$, A is a *rectangular* matrix; when $m = n$, A is $m \times m$ or $n \times n$, and A is a square matrix, having the same number of rows or columns.

1.1 Rows, columns and vectors

Clearly, there is no reason why m or n cannot equal 1: so, an $m \times n$ matrix with $n = 1$, i.e. with one column, is usually called a column vector. Similarly, a matrix with one row is a row vector.

There are a lot of advantages to thinking of matrices as collections of row or column vectors, as we shall see. As an example, define the 2×1 column vectors

$$\mathbf{a} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

and arrange as the columns of the 2×2 matrix

$$A = \begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}. \tag{1.1}$$

In general, a column vector $\mathbf{x}$ with n elements can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

What happens when both m and n are equal to 1? Then, A is a 1×1 matrix, but it is also considered to be a real number, or *scalar* in the language of linear algebra:

$$A = [a_{11}] = a_{11}.$$

This is perhaps a little odd, but turns out to be a useful convention in a number of situations.

1.2 Transposition of vectors

The *rows* of the matrix A in equation (1.1) can be seen as elements of column vectors, say

$$\mathbf{c} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$$

This representation of row vectors as column vectors is a bit clumsy, so some transformation which converts a column vector into a row vector, and vice versa would be useful. The process of converting a column vector $\mathbf{a}$ into a row vector is called *transposition,* and the transposed version of $\mathbf{c}$ is denoted

$$\mathbf{c}^T = \begin{bmatrix} 6 & 2 \end{bmatrix},$$

the ^{T} superscript denoting transposition. In practice, a prime, $'$, is used instead of ^{T}. However, whilst the prime is much simpler to write than the ^{T} sign, it is also much easier to lose track of in writing out long or complicated expressions. So, it is best initially to use ^{T} to denote transposition rather than the prime $'$.

A can then be written via its rows as

$$A = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{d}^T \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}.$$

The same ideas can be applied to the matrices P and Q.

2 Operations with matrices

2.1 Addition, subtraction and scalar multiplication

For vectors, addition and subtraction are defined only for vectors of the same dimensions: if

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \mathbf{x} - \mathbf{y} = \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{bmatrix}.$$

Clearly, the addition or subtraction operation is *elementwise.* If $\mathbf{x}$ and $\mathbf{y}$ have different dimensions, there will be some elements left over once all the elements of the smaller dimensioned vector have been used up.

Another operation is *scalar multiplication*: if λ is a real number or scalar, the product $\lambda\mathbf{x}$ is defined as

$$\lambda\mathbf{x} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix},$$

so that every element of $\mathbf{x}$ is multiplied by the same scalar λ.

The two types of operation can be combined into the *linear combination* of vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\lambda\mathbf{x} + \mu\mathbf{y} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix} + \begin{bmatrix} \mu y_1 \\ \vdots \\ \mu y_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 + \mu y_1 \\ \vdots \\ \lambda x_n + \mu y_n \end{bmatrix}.$$

Equally, one can define the linear combination of vectors $\mathbf{x}, \mathbf{y}, \ldots, \mathbf{z}$ by scalars $\lambda, \mu, \ldots, \nu$ as

$$\lambda\mathbf{x} + \mu\mathbf{y} + \ldots + \nu\mathbf{z}$$

with typical element

$$\lambda x_i + \mu y_i + \ldots + \nu z_i,$$

provided that all the vectors have the same dimension.

For matrices, these ideas carry over immediately: apply to each column of the matrices involved. For example, if $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{bmatrix}$, both $m \times n$, then addition and subtraction are defined elementwise, as for vectors:

$$A + B = \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \ldots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix} = \left\| a_{ij} + b_{ij} \right\|,$$
$$A - B = \begin{bmatrix} \mathbf{a}_1 - \mathbf{b}_1 & \ldots & \mathbf{a}_n - \mathbf{b}_n \end{bmatrix} = \left\| a_{ij} - b_{ij} \right\|.$$

Scalar multiplication of A by λ involves multiplying every column vector of A by λ, and therefore multiplying every element of A by λ :

$$\lambda A = \begin{bmatrix} \lambda\mathbf{a}_1 & \ldots & \lambda\mathbf{a}_n \end{bmatrix} = \left\| \lambda a_{ij} \right\|.$$

With the same idea for B, the linear combination of A and B by λ and μ is

$$\lambda A + \mu B = \left[\ \lambda \mathbf{a}_1 + \mu \mathbf{b}_1 \quad \ldots \quad \lambda \mathbf{a}_n + \mu \mathbf{b}_n \ \right] = \left\| \lambda a_{ij} + \mu b_{ij} \right\|.$$

For example, consider the matrices

$$A = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

with $\lambda = 1, \mu = -2$: then

$$\begin{aligned} \lambda A + \mu B &= A - 2B \\ &= \begin{bmatrix} 4 & 0 \\ 1 & 7 \end{bmatrix}. \end{aligned}$$

2.2 Matrix - vector products

2.2.1 Inner product

The simplest form of a matrix vector product is the case where A consists of one row, so that A is $1 \times n$:

$$A = \mathbf{a}^T = \left[\ a_1 \quad \ldots \quad a_n \ \right].$$

If $\mathbf{x}$ is an $n \times 1$ vector,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

the product $A\mathbf{x} = \mathbf{a}^T \mathbf{x}$ is called the *inner product* and is defined as

$$\mathbf{a}^T \mathbf{x} = a_1 x_1 + \ldots + a_n x_n.$$

One can see that the definition amounts to multiplying corresponding elements in $\mathbf{a}$ and $\mathbf{x}$, and adding up the resultant products. Writing

$$\mathbf{a}^T \mathbf{x} = \left[\ a_1 \quad \ldots \quad a_n \ \right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_1 x_1 + \ldots + a_n x_n$$

motivates the familiar description of the *across and down rule* for this product: *across and down* is the 'multiply corresponding elements' part of the definition.

Notice that the result of the inner product is a real number, for example,

$$\mathbf{c}^T = \left[\ 6 \quad 2 \ \right], \quad \mathbf{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \quad \mathbf{c}^T \mathbf{x} = \left[\ 6 \quad 2 \ \right] \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 36 + 6 = 42.$$

In general, in the product $\mathbf{a}^T \mathbf{x}$, $\mathbf{a}$ and $\mathbf{x}$ must have the same number of elements, n say, for the product to be defined. If $\mathbf{a}$ and $\mathbf{x}$ had different numbers of elements, there would be some elements of $\mathbf{a}$ or $\mathbf{x}$ left over or not used in the product: e.g.

$$\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}.$$

When the inner product of two vectors is defined, the vectors are said to be *conformable*.

2.3 Orthogonality

Two vectors $\mathbf{x}$ and $\mathbf{y}$ with the property that $\mathbf{x}^T\mathbf{y} = 0$ are said to be orthogonal to each other. For example, if

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

it is clear that $\mathbf{x}^T\mathbf{y} = 0$. This seems a rather innocuous definition, and yet the idea of orthogonality turns out to be extremely important in econometrics.

If $\mathbf{x}$ and $\mathbf{y}$ are thought of as points in R^2, and arrows are drawn from the origin to $\mathbf{x}$ and to $\mathbf{y}$, then the two arrows are perpendicular to each other - see Figure 2.1. If $\mathbf{y}$ were defined as

$$\mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

the position of the $\mathbf{y}$ vector and the corresponding arrow would change, but the perpendicularity property would still hold.
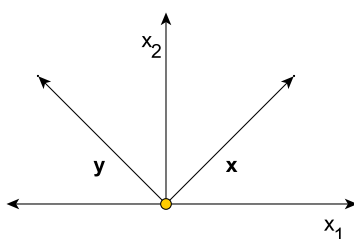

Figure 2.1: Orthogonality of the vectors $\mathbf{x}$ and $\mathbf{y}$.

2.3.1 Matrix - vector products

Since the matrix

$$A = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}$$

has two rows, now denoted $\boldsymbol{\alpha}_1^T$ and $\boldsymbol{\alpha}_2^T$, there are two possible inner products with the vector

$$\mathbf{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}:$$

$$\boldsymbol{\alpha}_1^T\mathbf{x} = 42, \quad \boldsymbol{\alpha}_2^T\mathbf{x} = 33.$$

Assembling the two inner product values into a 2×1 vector defines the product of the matrix A with the vector $\mathbf{x}$:

$$A\mathbf{x} = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1^T\mathbf{x} \\ \boldsymbol{\alpha}_2^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} 42 \\ 33 \end{bmatrix}.$$

Focussing only on the part

$$\begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 42 \\ 33 \end{bmatrix},$$

one can see that each element of $A\mathbf{x}$ is obtained from an *across and down* argument.

Sometimes this product is described as forming a *linear combination* of the columns of A using the scalar elements in $\mathbf{x}$:

$$A\mathbf{x} = 6 \begin{bmatrix} 6 \\ 3 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

More generally, if

$$A = \begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \lambda \\ \mu \end{bmatrix},$$
$$A\mathbf{x} = \lambda \mathbf{a} + \mu \mathbf{b}.$$

The general version of these ideas for an $m \times n$ matrix A,

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix}.$$

is straightforward. If $\mathbf{x}$ is an $n \times 1$ vector, then the vector $A\mathbf{x}$ is, by the *across and down* rule,

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ a_{21}x_1 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j}x_j \\ \sum\limits_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{mj}x_j \end{bmatrix},$$

(2.1)

so that the typical element, the ith, is $\sum\limits_{j=1}^{n} a_{ij}x_j$. Equally, $A\mathbf{x}$ is the linear combination

$$\mathbf{a}_1 x_1 + \ldots + \mathbf{a}_n x_n$$

of the columns of A.

2.4 Matrix - matrix products

Suppose that A is $m \times n$, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and B is $n \times r$, with columns $\mathbf{b}_1, \ldots, \mathbf{b}_r$. Clearly, each product $A\mathbf{b}_1, \ldots, A\mathbf{b}_r$ exists, and is $m \times 1$. These products can be arranged as the columns of a matrix as

$$\begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_r \end{bmatrix}$$

and this matrix is *defined* to be the product C of the matrices A and B:

$$C = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_r \end{bmatrix} = AB.$$

By construction, this must be an $m \times r$ matrix, since each column is $m \times 1$ and there are r columns.

This is not the usual presentation of the definition of the product of two matrices, which relies on the *across and down rule* mentioned earlier, and focusses on the elements of each matrix A and B. Set

$$
\begin{aligned}
B &= \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_r \end{bmatrix} \quad \text{(by columns)} \\
&= \|b_{ik}\|, \quad i = 1, \ldots, n, k = 1, \ldots, r \quad \text{(typical element)} \\
&= \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1r} \\ b_{21} & b_{22} & \ldots & b_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{nr} \end{bmatrix} \quad \text{(the array)}
\end{aligned}
$$

What does the typical element of the $m \times r$ matrix C look like? Start with the kth column of C, which is $A\mathbf{b}_k$. The ith element in $A\mathbf{b}_k$ is, from equation (2.1), the inner product of the elements of the ith row in A,

$$
\begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix},
$$

with the elements of $\mathbf{b}_k$, so that the inner product is

$$
a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.
$$

So, the ikth element of C is

$$
c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.
$$

We can see this arising from an *across and down* calculation by writing

$$
\begin{aligned}
C &= AB \quad (2.2) \\
&= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1k} & \ldots & b_{1r} \\ b_{21} & b_{22} & \ldots & b_{2k} & \ldots & b_{2r} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{nk} & \ldots & b_{nr} \end{bmatrix} \quad (2.3) \\
&= \left\| \sum_{j=1}^{n} a_{ij}b_{jk} \right\|. \quad (2.4)
\end{aligned}
$$

These ideas are simple, but a little tedious. Numerical examples are equally tedious! As an example, using

$$
A = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix},
$$

we can find the matrix B such that

1. the first column of AB adds together the columns of A,

2. the second column is the difference of the first and second columns of A,

3. the third column is $2\times$ the first column of A,

4. the fourth column is zero.

It is easy to check that B is

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

and that

$$\begin{aligned} C & = AB \\ & = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix} \\ & = \begin{bmatrix} 8 & 4 & 12 & 0 \\ 8 & -2 & 6 & 0 \end{bmatrix}. \end{aligned}$$

Arithmetic calculations of matrix products almost always use the elementwise across and down formula. However, there are many situations in econometrics where algebraic rather than arithmetic arguments are required. In these cases, the viewpoint of matrix multiplication as linear combinations of columns is much more powerful.

Clearly one can give many more examples of different dimensions and complexities - but the same basic rules apply. To multiply two matrices A and B together, the number of columns in A must match the number of rows in B - this is *conformability* in action again. The resulting product will have number of rows equal to the number in A and number of columns equal to the number in B.

If this conformability rule does not hold, then the product of A and B is not defined.

2.5 Matlab

One should also say that as the dimensions of the matrices increases, so the tediousness of the calculations increase. The solution to this for numerical calculation is to appeal to the computer. Programs like Matlab and Excel (and a number of others, some of them free) resolve this difficulty easily.

In Matlab, symbols for row or column vectors do not need any particular differentiation: they are distinguished by how they are defined. For example, the following Matlab commands define `rowvec` as a 1×4 vector, and `colvec` as a 4×1 vector, then display the contents of these variables, and do a calculation:

```
>> rowvec = [1 2 3 4];
>> colvec = [1;2;3;4];
>> rowvec
rowvec =
1 2 3 4
>> colvec
colvec =
1 2 3 4
>> rowvec*colvec
ans =
30
```

So, the semi-colon indicates the end of a row in a matrix or vector; it can be replaced by a carriage return. Notice the difference in how a row vector and a column vector is defined. One can see that the product `rowvec*colvec` is well defined, just because `rowvec` is a 1×4 vector, and `colvec` is a 4×1 vector.

Matlab also allows elementwise multiplication of two vectors using the `.*` operator: if

$$\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right], \quad \mathbf{y} = \left[\begin{array}{c} y_1 \\ y_2 \end{array} \right],$$

then

$$\mathbf{x}.*\mathbf{y} = \left[\begin{array}{c} x_1 y_1 \\ x_2 y_2 \end{array} \right]$$

and one can see that the inner product of $\mathbf{x}$ and $\mathbf{y}$ can be obtained as the sum of the elements of $\mathbf{x}$ and $\mathbf{y}$. In Matlab, this would be obtained as

$$\text{sum} \left(\mathbf{x}.*\mathbf{y} \right).$$

In the example above, this calculation fails since `rowvec` is a 1×4 vector, and `colvec` is a 4×1 vector:

```
>> sum(rowvec .* colvec) ???
Error using ==> times Matrix dimensions must agree.
```

For this to work, `rowvec` would have to be transposed as `rowvec'`, so that transposition in Matlab is very natural.

Allowing for such difficulties, matrix multiplication in Matlab is very simple:

```
>> A = [6 2; 3 5];
>> B = [1 1 2 0;1 -1 0 0];
>> C = A * B;
>> disp(C)
 8 4 1
 2 0 8
-2 6 0
```

Notice how the matrices are defined here through their rows. The `disp()` command displays the contents of the object referred to.

It is less natural in Matlab to define matrices by columns - a typical example of how mathematics and computing have conflicts of notation. However, once columns $\mathbf{a}$ and $\mathbf{b}$ have been defined, the concatenation operation $\left[\begin{array}{cc} \mathbf{a} & \mathbf{b} \end{array} \right]$ collects the columns into a matrix:

```
>> a = [6;2];
>> b = [3;5];
>> C = [a b];
>> disp(C)
6 3
2 5
```

Notice that the `disp(C)` command does not label the result that is printed out. Simply typing `C` would preface the output by `C =`.

2.6 Pre and Post Multiplication

If $C = AB$, as above, say that B is *pre-multiplied* by A to get C, and that A is *post-multiplied* by B to get C.

This distinction between *pre* and *post* multiplication is important, in the following sense. Suppose that A and B are matrices such that the products AB and BA are both defined. If A is $m \times n$, B must have n rows for AB to be defined. For BA to be defined, B must have m columns to match the m rows in A. So, AB and BA are both defined if A is $m \times n$ and B is $n \times m$.

Even when both products are defined, there is no reason for the two products coincide. The first thing to notice is that AB is a square, $m \times m$, matrix, whilst BA is a square, $n \times n$, matrix. Different sized matrices cannot be equal. To illustrate, use the matrices

$$B_2 = \begin{bmatrix} 6 & -3 \\ 2 & 5 \\ -3 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 6 & 2 & -3 \\ 3 & 5 & -1 \end{bmatrix}:$$

$$B_2 C = \begin{bmatrix} 6 & -3 \\ 2 & 5 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 & -3 \\ 3 & 5 & -1 \end{bmatrix} = \begin{bmatrix} 27 & -3 & -15 \\ 27 & 29 & -11 \\ -15 & -1 & 8 \end{bmatrix},$$

$$CB_2 = \begin{bmatrix} 6 & 2 & -3 \\ 3 & 5 & -1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 2 & 5 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 49 & -11 \\ 31 & 15 \end{bmatrix}.$$

Even when $m = n$, so that AB and BA are both $m \times m$ matrices, the products can differ: for example,

$$A = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

$$AB = \begin{bmatrix} 8 & 4 \\ 8 & -2 \end{bmatrix}, \quad BA = \begin{bmatrix} 9 & 7 \\ 3 & -3 \end{bmatrix}.$$

In cases where $AB = BA$, the matrices A and B are said to *commute*.

2.7 Transposition

A column vector $\mathbf{x}$ can be converted to a row vector $\mathbf{x}^T$ by transposition:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{x}^T = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}.$$

Transposing $\mathbf{x}^T$ as $(\mathbf{x}^T)^T$ reproduces the original vector $\mathbf{x}$. How do these ideas carry over to matrices?

If the $m \times n$ matrix A can be written as $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$, the transpose of A, A^T, is defined as the matrix whose *rows* are $\mathbf{a}_i^T$:

$$A^T = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}.$$

In terms of elements, if

$$
\mathbf{a}_i = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix}
$$

then

$$
A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad A^T = \begin{bmatrix} a_{11} & \cdots & a_{i1} & \cdots & a_{m1} \\ a_{12} & \cdots & a_{i2} & \cdots & a_{m2} \\ \vdots & & \vdots & & \vdots \\ a_{1n} & \cdots & a_{in} & \cdots & a_{mn} \end{bmatrix}.
$$

One can see that the first column of A has now become the first row of A^T. Notice too that A^T is an $n \times m$ matrix if A is an $m \times n$ matrix.

Transposing A^T takes the first column of A^T and writes it as a row, which coincides with the first row of A. The same argument applies to the other columns of A^T, so that

$$
\left(A^T \right)^T = A.
$$

2.7.1 The product rule for transposition

This states that if $C = AB$, then $C^T = B^T A^T$.

How to see this? Consider the following example,

$$
A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{bmatrix}
$$

where

$$
c_{23} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} = \sum_{k=1}^{3} a_{2k}b_{k3}. \tag{2.5}
$$

One can see that

$$
B^T A^T = \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \\ b_{14} & b_{24} & b_{34} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}
$$

and that the $(3,2)$ element of this product is actually c_{23} :

$$
b_{13}a_{21} + b_{23}a_{22} + b_{33}a_{23} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} = c_{23}.
$$

In summation notation, we see that from $B^T A^T$

$$
c_{23} = \sum_{k=1}^{3} b_{k3}a_{2k},
$$

where the position of the index of summation is due to the transposition. So, in summation notation, the calculation of c_{23} from $B^T A^T$ equals that from equation (2.5).

More generally, the (i, j) element of AB,

$$\sum_{k=1}^{3} a_{ik} b_{kj}$$

is the (j, i) element of $B^T A^T$. But this means that $B^T A^T$ must be the transpose of AB, since the elements in the ith row of AB are being written in the ith column of $B^T A^T$.

This *Product Rule for Transposition* can be applied again to find the transpose $\left(C^T\right)^T$ of C^T :

$$\left(C^T\right)^T = \left(B^T A^T\right)^T = \left(A^T\right)^T \left(B^T\right)^T = AB = C.$$

3 Special Types of Matrix

3.1 The zero matrix

The most obvious special type of matrix is one whose elements are all zeros. In typical element notation, the zero matrix is

$$0 = \|0\| .$$

Since there is no indexing on the elements, it is not obvious what the dimension of this matrix is, Sometimes one writes 0_{mn} to indicate a zero matrix of dimension $m \times n$. The same ideas apply to vectors whose elements are all zero.

The effect of the zero matrix in any product that is defined is simple:

$$0A = 0, \quad B0 = 0.$$

This is easy to check using the across and down rule.

3.2 The identity or unit matrix

Vectors of the form

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{in 2 dimensions}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{in 3 dimensions}$$

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad \text{in } n \text{ dimensions}$$

are called coordinate vectors. They are often given a characteristic notation, $\mathbf{e}_1, \dots, \mathbf{e}_n$, in n dimensions. When arranged as columns of a matrix in the natural order, $\mathbf{e}_1, \dots, \mathbf{e}_n$,

a matrix with a characteristic pattern elements emerges, with a special notation:

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n.$$

The *diagonal* of this matrix is where the 1 elements are located, and every other element is zero.

Consider the effect of I_2 on the matrix

$$A = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix}$$

by both pre and post multiplication:

$$I_2 A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} = A,$$

$$A I_2 = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} = A,$$

as is easily checked by the across and down rule.

Because any matrix is left unchanged by pre or post multiplication by an appropriately dimensioned I_n, I_n is called an *identity matrix of dimension n*. Sometimes it is called a *unit matrix of dimension n*. Notice that I_n is necessarily a square matrix.

3.3 Diagonal matrices

The identity matrix is an example of a diagonal matrix, a matrix whose elements are all zero except for those on the diagonal. Usually diagonal matrices are taken to be square, for example

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

They also produce characteristic effects when pre or post multiplying another matrix.

Consider the diagonal matrix

$$B = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

and the products AB, BA for A as defined in the previous section:

$$AB = \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 12 & -4 \\ 6 & -10 \end{bmatrix},$$

$$BA = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 12 & 4 \\ -6 & -10 \end{bmatrix}.$$

Comparing the results, we can deduce that post multiplication by a diagonal matrix multiplies each column of A by the corresponding diagonal element, whereas pre multiplication multiplies each row by the corresponding diagonal element.

3.4 Symmetric matrices

Symmetric matrices are matrices having the property that $A = A^T$. Notice that such matrices must be square, since if A is $m \times n$, A^T is $n \times m$, and to have equality of A and A^T, they must have the same dimension, so that $m = n$ is required.

Suppose that A is a 3×3 symmetric matrix, with typical element a_{ij} :

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

so that

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

Equality of matrices is defined as equality of all elements. This is fine on the diagonals, since A and A^T have the same diagonal elements. For the off diagonal elements, we end up with the requirements

$$a_{12} = a_{21}, \quad a_{13} = a_{31}, \quad a_{23} = a_{32}$$

or more generally

$$a_{ij} = a_{ji} \quad \text{for } i \neq j.$$

The effect of this conclusion is that in a symmetric matrix, the 'triangle' of above diagonal elements coincides with the triangle of below diagonal elements. It is as if the upper triangle is folded over the diagonal to become the lower triangle.

A simple example is

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

A more complicated example uses the 2×3 matrix C

$$C = \begin{bmatrix} 6 & 2 & -3 \\ 3 & 5 & -1 \end{bmatrix}$$

and calculates the 3×3 matrix

$$C^T C = \begin{bmatrix} 6 & 3 \\ 2 & 5 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} 6 & 2 & -3 \\ 3 & 5 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 45 & 27 & -21 \\ 27 & 29 & -11 \\ -21 & -11 & 10 \end{bmatrix},$$

which is clearly symmetric.

This illustrates the general proposition that if A is an $m \times n$ matrix, the product $A^T A$ is a symmetric $n \times n$ matrix. Proof? Compute the transpose of $A^T A$ using the product rule for transposition:

$$\left(A^T A\right)^T = A^T \left(A^T\right)^T = A^T A.$$

Since $A^T A$ is equal to its transpose, it must be a symmetric matrix. Such symmetric matrices appear frequently in econometrics.

It should be clear that diagonal matrices are symmetric, since all their off-diagonal elements are equal (zero), and thence the identity matrix I_n is also symmetric.

3.5 The outer product

The inner product of two $n \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}^T \mathbf{y}$, is automatically a 1×1 quantity, a scalar, although it can be interpreted as a 1×1 matrix, a matrix with a single element.

Suppose one considered the product of $\mathbf{x}$ with $\mathbf{x}^T$. Is this defined? If A is $m \times n$ and B is $n \times r$, then the product AB is $m \times r$. Applying this logic to $\mathbf{x}\mathbf{x}^T$, this is $(n \times 1)(1 \times n)$, so the resulting product *is* defined, and is an $n \times n$ *matrix* - the *outer product* of $\mathbf{x}$ and $\mathbf{x}^T$, the word 'outer' being used to distinguish from the inner product.

How does the across and down rule work here? Suppose that

$$\mathbf{x} = \left[\begin{array}{c} 6 \\ 3 \end{array} \right].$$

Then,

$$\mathbf{x}\mathbf{x}^T = \left[\begin{array}{c} 6 \\ 3 \end{array} \right] \left[\begin{array}{cc} 6 & 3 \end{array} \right].$$

Here, there is 1 element in row one of the 'matrix' $\mathbf{x}$, and 1 element in column one of the matrix $\mathbf{x}^T$, so the across and down rule still works - it is just that there is only one product per row and column combination. So,

$$\mathbf{x}\mathbf{x}^T = \left[\begin{array}{cc} 36 & 18 \\ 18 & 9 \end{array} \right],$$

and it is obvious from this that $\mathbf{x}\mathbf{x}^T$ is a symmetric matrix.

One can see that this outer product need not be restricted to vectors of the same dimension. If $\mathbf{x}$ is $n \times 1$ and $\mathbf{y}$ is $m \times 1$, then

$$\mathbf{x}\mathbf{y}^T = \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right] \left[\begin{array}{ccc} y_1 & \cdots & y_m \end{array} \right] = \left[\begin{array}{cccc} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ & & & \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{array} \right].$$

So, $\mathbf{x}\mathbf{y}^T$ is $n \times m$, and consists of rows which are $\mathbf{y}^T$ multiplied by an element of the $\mathbf{x}$ vector.

Another interesting and useful example involves a vector with every element equal to 1,

$$\mathbf{1} = \left[\begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right].$$

Sometimes this is written as $\mathbf{1}_n$ to indicate an $n \times 1$ vector, and is called the *sum vector*. Why? Consider the impact of $\mathbf{1}_2$ on the 2×1 vector $\mathbf{x}$ used above:

$$\mathbf{1}_2^T \mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 9,$$

i.e. an inner product of $\mathbf{x}$ with the sum vector is the sum of the elements of $\mathbf{x}$. Dividing through by the number of elements in $\mathbf{x}$ produces the average of the elements of $\mathbf{x}$ - i.e. the 'sample mean' of the elements of $\mathbf{x}$.

The outer product of $\mathbf{x}$ with $\mathbf{1}_2$ is also interesting:

$$\begin{aligned} \mathbf{1}_2 \mathbf{x}^T &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 6 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 6 & 3 \end{bmatrix}, \\ \mathbf{x} \mathbf{1}_2^T &= \begin{bmatrix} 6 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 3 & 3 \end{bmatrix}, \end{aligned}$$

showing that pre multiplication of an $\mathbf{x}^T$ by $\mathbf{1}$ repeats $\mathbf{x}^T$ as rows of the product, whilst post multiplication of $\mathbf{x}$ by $\mathbf{1}^T$ repeats $\mathbf{x}$ as the columns of the product.

Finally,

$$\mathbf{1}_n \mathbf{1}_n^T = \begin{bmatrix} 1 & \dots & 1 \\ 1 & \dots & 1 \\ 1 & \dots & 1 \end{bmatrix},$$

an $n \times n$ matrix with every element equal to 1. This type of matrix also appears in econometrics!

3.6 Triangular matrices

A square *lower triangular* matrix has all elements above the main diagonal equal to zero, whilst a square *upper triangular* matrix has all elements below the main diagonal equal to zero. A simple example of a lower triangular matrix is

$$A = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Clearly, for this matrix, A^T is an upper triangular matrix.

One can adapt the definition to rectangular matrices: for example, if two arbitrary rows are added to A, so that it becomes 5×3, it would still be considered lower triangular. Equally, if, for example, the third column of A above is removed, A is still considered lower triangular.

Often, we use *unit* triangular matrices, where the diagonal elements are all equal to 1 : e.g.

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.1}$$

3.7 Partitioned matrices

Sometimes, especially with big matrices, it is useful to organise the elements of the matrix into components which are themselves matrices, for example,

$$B = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 8 & 3 & 0 & 0 \\ 0 & 0 & 7 & 4 \\ 0 & 0 & 6 & 5 \end{bmatrix}$$

Here it would be reasonable to write

$$B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix},$$

where B_{ii}, $i = 1, 2$, represent 2×2 matrices. B is an example of a *partitioned matrix*: that is, an $m \times n$ matrix A say,

$$A = \|a_{ij}\|,$$

where the elements of A are organised into *sub-matrices*. An example might be

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{23} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}, \tag{3.2}$$

where the *sub - matrices* in the first row block have r rows, and therefore $m - r$ rows in the second row block. The column blocks might be defined by (for example) 3 columns in the first column block, 4 in the second and $n - 7$ in the third column block.

Another simple example might be

$$A = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix},$$

where A and therefore A_1, A_2, A_3 have m rows, A_1 has n_1 columns, A_2 has n_2 columns, A_3 has n_3 columns. The *subvectors* in $\mathbf{x}$ must have n_1, n_2 and n_3 rows respectively, for the product $A\mathbf{x}$ to exist.

Suppose that $n_1 + n_2 + n_3 = n$, so that A is $m \times n$. The ith element of $A\mathbf{x}$ is

$$\sum_{i=1}^{n} a_{ij} x_j$$

but the summation can be broken up into the first n_1 terms,

$$\sum_{i=1}^{n_1} a_{ij} x_j,$$

the next n_2 terms,

$$\sum_{i=n_1+1}^{n_1+n_2} a_{ij} x_j,$$

and the next n_3 terms,

$$\sum_{i=n_1+n_2+1}^{n} a_{ij} x_j.$$

The point about the use of partitioned matrices is that the product $A\mathbf{x}$ can be represented as

$$A\mathbf{x} = A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A\mathbf{x}_3$$

by applying the across and down rule to the submatrices and the subvectors, a much simpler representation than the use of summations.

Each of the components is a conformable matrix-vector product: this is essential in any use of partitioned matrices to represent some matrix product. For example, using A from equation (3.2) and B as

$$B = \begin{bmatrix} B_{11} \\ B_{21} \\ B_{31} \end{bmatrix},$$

it is easy to write

$$
\begin{aligned}
AB &= \begin{bmatrix} A_{11} & A_{12} & A_{23} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{21} \\ B_{31} \end{bmatrix} \\
&= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31} \\ A_{21}B_{11} + A_{22}B_{21} + A_{23}B_{31} \end{bmatrix}.
\end{aligned}
$$

But, what are the row dimensions for the submatrices in B? What are the possible column dimensions for the submatrices in B?

4 Matrices, vectors and econometrics

The data on weights and heights for 12 students in the data matrix

$$D = \begin{bmatrix} 155 & 70 \\ 150 & 63 \\ 180 & 72 \\ 135 & 60 \\ 156 & 66 \\ 168 & 70 \\ 178 & 74 \\ 160 & 65 \\ 132 & 62 \\ 145 & 67 \\ 139 & 65 \\ 152 & 68 \end{bmatrix}$$

would seem to be ideally suited for fitting a two variable regression model

$$y_i = \alpha + \beta x_i + u_i, \quad i = 1, \ldots, 12.$$

Here, the first column of D contains all the weight data, the data on the dependent variable y_i, and so should be labelled $\mathbf{y}$. The second column of D contains all the data on the explanatory variable height, in the vector $\mathbf{x}$ say, so that

$$D = \begin{bmatrix} \mathbf{y} & \mathbf{x} \end{bmatrix}.$$

If we define a 12×1 vector with every element 1,

$$\mathbf{1}_{12} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

and a 12×1 vector $\mathbf{u}$ to contain the error terms,

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_{12} \end{bmatrix}$$

the regression model can be written in terms of the three data vectors $\mathbf{y}, \mathbf{1}_{12}$ and $\mathbf{x}$ as

$$\mathbf{y} = \mathbf{1}_{12}\alpha + \mathbf{x}\beta + \mathbf{u}.$$

To see this, think of the ith elements of the vectors on the left and right hand sides.

The standard next step is then to combine the data vectors for the explanatory variables into a matrix

$$X = \begin{bmatrix} \mathbf{1}_{12} & \mathbf{x} \end{bmatrix},$$

and then define a 2×1 vector δ to contain the parameters α, β as

$$\delta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

to give the data matrix representation of the regression model as

$$\mathbf{y} = X\delta + \mathbf{u}.$$

For the purposes of developing the theory of regression, this is the most convenient form of the regression model. It can represent regression models with any number of explanatory variables, and thus any number of parameters. The obvious point is that a knowledge of vector and matrix operations is needed to use and understand this form.

We shall see later that there are two particular matrix and vector quantities associated with a regression model. The first is the matrix $X^T X$, and the second the vector $X^T \mathbf{y}$. The following Matlab code snippet provides the numerical values of these quantities for the weight data:

```
>> dset = load('weights.mat');
>> xtx = dset.X' * dset.X;
>> xty = dset.X' * dset.y;
>> disp(xtx)
   12    802
  802  53792
>> disp(xty)
    1850
  124258
```

Hand calculation is of course possible, but not recommended.