APPENDIX 1: An Introduction to Matrix Algebra and Linear Models

Julius van der Werf and Brian Kinghorn
University of New England, Armidale, Australia

> Matrices take a lot of the tedium out of both presenting algebra and calculating results. They are widely used in scientific animal breeding, particularly in selection indices and BLUP.

A matrix is a rectangular array with dimensions Rows x Columns.

For example:

$$A = \begin{pmatrix} 4 & 6 \\ 7 & 3 \\ 2 & 1 \end{pmatrix}$$

is a 3 (rows) x 2 (columns) matrix. Element a_{ij} has value 6. Note the convention of using a small letter for elements, and subscripts denoting row and column, in that order.

$$B = \begin{pmatrix} 4 \\ 7 \\ 9 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 8 & 3 & 14 \end{pmatrix}$$

are matrices which can be referred to as vectors. B is a column vector and C is a row vector, both of length 3.

Matrices of equal dimensions can be added and subtracted

Weaning weight, Kg	Blue Angus	Leanford	Meatmaker
High Nutrition	120	140	150
Low Nutrition	90	100	105

Yearling weight, Kg	Blue Angus	Leanford	Meatmaker
High Nutrition	260	290	320
Low Nutrition	220	250	275

These hypothetical data can be represented in Nutrition x Breed matrices. Note the special meaning that element locations have - they indicate breed and nutrition:

$$W = \begin{pmatrix} 120 & 140 & 150 \\ 90 & 100 & 105 \end{pmatrix} \quad Y = \begin{pmatrix} 260 & 290 & 320 \\ 220 & 250 & 275 \end{pmatrix}$$

$$G = \begin{pmatrix} 140 & 150 & 170 \\ 130 & 150 & 170 \end{pmatrix}$$

Note that the matrix of Growth (Y - W = G) is of the same dimension as the others, and is simply got by subtracting elements of W from corresponding elements of Y

Matrices can be multiplied by a scalar - a simple constant:

For example, to express G in Lbs rather that KG, multiply by 2.2:

$$G_{Lbs} = 2.2 \begin{pmatrix} 140 & 150 & 170 \\ 130 & 150 & 170 \end{pmatrix} = \begin{pmatrix} 308 & 330 & 374 \\ 286 & 330 & 374 \end{pmatrix}$$

Matrix multiplication:

In this hypothetical example we have information to make two matrices:

M: a matrix of merit for breeds (X, Y and Z) by traits (Body weight and backfat).

P: a matrix of dollars per unit for traits (Body weight and Backfat) by markets (Domestic and Export).

And the product of these matrices will be:

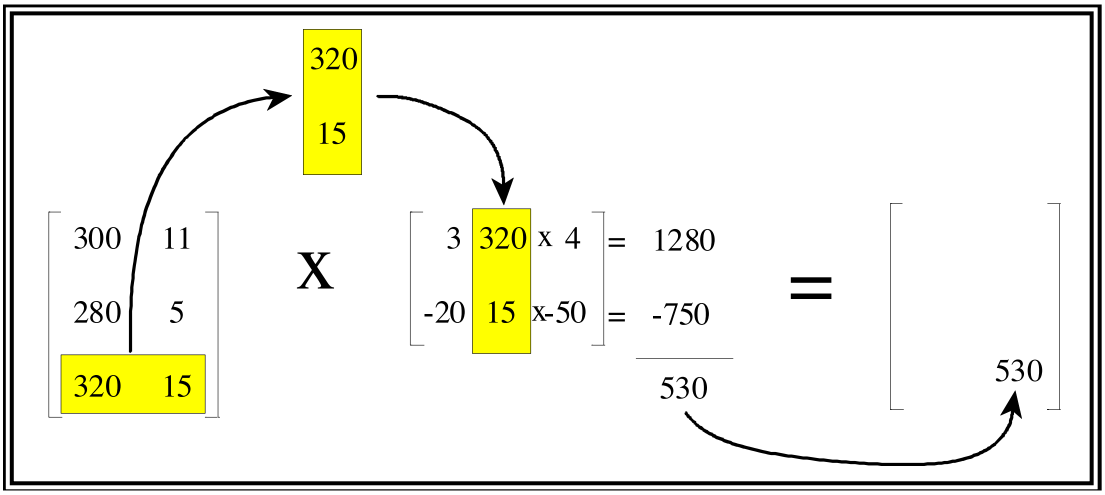
R: a matrix of dollars per head for Breeds (X, Y and Z) by markets (Domestic and Export).

	M		P		R
	MERIT MATRIX	**x**	**PRICE MATRIX**	**=**	**RETURNS MATRI**

	Wt.	Fat		Dom.	Exp.		Dom.	Exp.
Breed X	300	11	Wt.	3	4	Breed X	680	650
Breed Y	280	5	Fat	-20	-50	Breed Y	740	870
Breed Z	320	15				Breed Z	660	530

Breeds** x Traits* **x** *Traits **x Markets = ***Breeds x Markets***

First note that the number of columns of M (= traits, 2) must equal the number of rows of P (also traits, 2) in order to be able to multiply. The following shows calculation of $r_{3,2}$

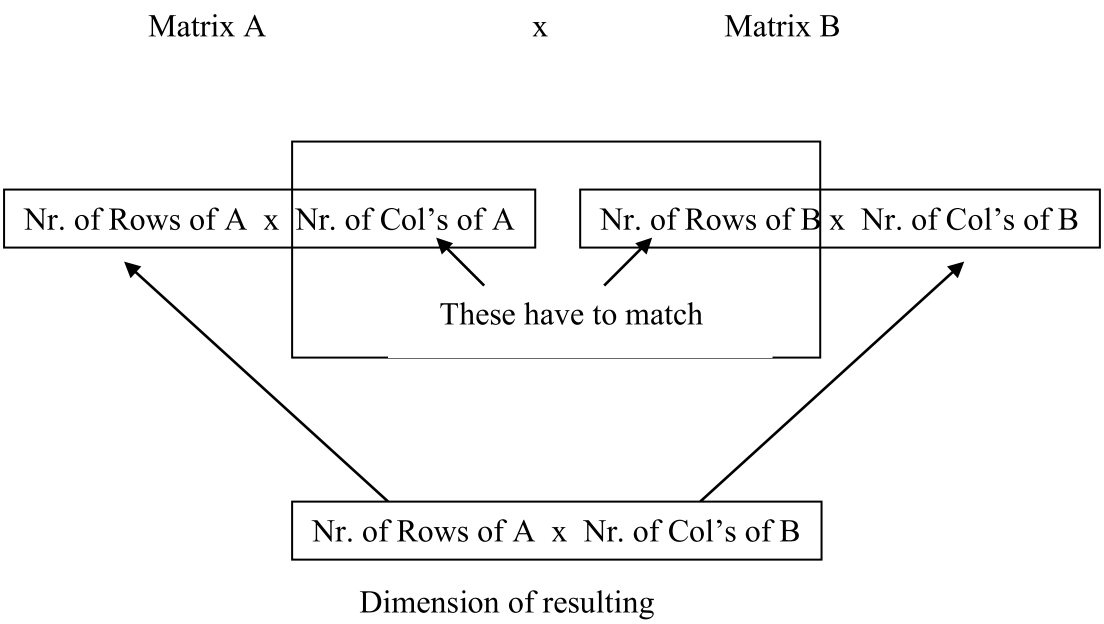


Matrix multiplication: Here is the same information in words:	
To calculate the elements of R:	The $(i,j)^{th}$ (row, column) of R is the sum of the products of the elements of the i^{th} row of M and the j^{th} column of P
For example, $r_{3,2}$:	$R_{3,2} = m_{3,1} \times p_{1,2} + m_{3,2} \times p_{2,2}$ 530 = 320 x 4 + 15 x -50

Another example:

$$\begin{pmatrix} 5 & 3 \\ 4 & 2 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \begin{pmatrix} 5x7 + 3x3 \\ 4x7 + 2x3 \\ 1x7 + 5x3 \end{pmatrix} = \begin{pmatrix} 44 \\ 34 \\ 22 \end{pmatrix}$$

For matrix multiplication to be legal, the first matrix must have as many columns as the second matrix has rows. This, of course, is also the requirement for multiplying a row vector by a column vector. The resulting matrix will have as many rows as the first matrix and as many columns as the second matrix. Because **A** has 2 rows and 3 columns while **B** has 3 rows and 2 columns, the matrix multiplication may legally proceed and the resulting matrix will have 2 rows and 2 columns.



Because of these requirements, matrix multiplication is usually not commutative.
That is, usually **AB** ≠ **BA**. And even if **AB** is a legal operation, there is no guarantee that **BA** will also be legal. For these reasons, the terms **premultiply** and **postmultiply** are often encountered in matrix algebra while they are seldom encountered in scalar algebra.

Whenever you propose matrix multiplications, make sure they 'fit'

One special case to be aware of is when a column vector is postmultiplied by a row vector. In this case, one simply follows the rules given above for the multiplication of two matrices. Note that the first matrix has one column and the second matrix has one row, so the matrix multiplication is legal. The resulting matrix will have as many rows as the first matrix (3) and as many columns as the second matrix (2).

$$\begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix} \begin{pmatrix} 3 & 4 \end{pmatrix} = \begin{pmatrix} 9 & 12 \\ 6 & 8 \\ 15 & 20 \end{pmatrix}$$

Similarly, multiplication of a matrix times a vector (or a vector times a matrix) will also conform to the multiplication of two matrices. For example,

$$\begin{pmatrix} 9 & 12 \\ 6 & 8 \\ 15 & 20 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}$$

is an illegal operation because the number of columns in the first matrix (2) does not match the number of rows in the second matrix (3).

There are a couple of examples that are worth looking let. Let us define the column vector **e**. By definition, the order of **e** is (*N,* 1). We can take the *inner product* of **e**, which is simply:

$$e'e = [e_1 \quad e_2 \cdots e_N] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = e_1e_1 + e_2e_2 + \cdots + e_Ne_N = \sum_{i=1}^{N} e_i^2$$

The inner product of a column vector with itself is simply equal to the sum of the square values of the vector, which is used quite often in the regression model. Geometrically, the square root of the inner product is the length of the vector.
There are couple of other vector products that are interesting to note. Let **i** denote an order (*N,* 1) vector of ones, and **x** denote an order (*N,* 1) vector of data. The following is an interesting quantity:

$$\frac{1}{N} i'x = \frac{1}{N}(x_1, \ldots, x_n) = \frac{1}{N}\sum_{i=1}^{n} x_i = \bar{x}$$ is the mean of all x_i

From this, it follows that: $i'x = \sum_{i=1}^{n} x_i$ is the sum of all x_i

Similarly, let **y** denote another (*N,* 1) vector of data. The following is also interesting:

$\mathbf{x'y} = x_1 y_1 + x_2 y_2 + \cdots + x_N y_N = \sum_{i=1}^{n} x_i y_i$ is the cross product of all $x_i y_i$

The identity matrix, I.

The number 1 is quite special, in that if you multiple any number by 1 that number retains its identity - it is not changed.

The same property holds for the identity matrix , which is a square matrix. There is not just one identity matrix, but one for each size, populated with zeros, except for the 'leading diagonal' (top left to bottom right) which contains one's.

You can check that the following are in fact true:

IA = A $\quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 4 & 6 & 7 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 7 \\ 3 & 2 & 1 \end{pmatrix}$

2x2   2x3   2x3

AI = A $\quad \begin{pmatrix} 4 & 6 & 7 \\ 3 & 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 7 \\ 3 & 2 & 1 \end{pmatrix}$

2x3   3x3   2x3

Note, that a scalar multiplied by an identity matrix becomes a diagonal matrix with the scalars on the diagonal.

Diagonal matrix

A diagonal matrix has only non-zero elements on its diagonal,

For example $\begin{pmatrix} 2.45 & 0 & 0 \\ 0 & 1.71 & 0 \\ 0 & 0 & 1.69 \end{pmatrix}$

Transpose - pivot the matrix about the top left element

$b = \begin{pmatrix} x \\ y \end{pmatrix}$ gives $b' = \begin{pmatrix} x & y \end{pmatrix}$ "b transpose"

The transpose of a matrix is denoted by a prime (′): **A'** or a superscript t or T ($\mathbf{A}^t$ or $\mathbf{A}^T$).

$$\begin{pmatrix} 4 & 6 \\ 7 & 3 \\ 2 & 1 \end{pmatrix}' = \begin{pmatrix} 4 & 7 & 2 \\ 6 & 3 & 1 \end{pmatrix}$$

Note that $a'_{i,j} = a_{j,i}$

The transpose of a product takes an interesting form: $(AB)' = B'A'$

Symmetrical matrix

A matrix is symmetrical if A = A'. A symmetrical matrix has to be also a squared matrix (equal numbers of rows and columns)

Matrix inversion **(for reference only)**

Scalar: $X^{-1} = {}^{1}/_{x}$

Only square matrices can be inverted. We do this in order to achieve matrix division – we just multiply by the reciprocal, or inverse! Just as 20/5 = 4, we have 20 x $5^{-1} = 4$

The inverse of a matrix is denoted by the superscript "-1"

Inverse of a 2 x 2 Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

For matrices larger that 2 x 2, inversion is quite tedious, and best left to a computer!

Exercise: Just as $X \cdot {}^{1}/_{x} = 1$ show that for matrices, $X X^{-1} = I$, the identity matrix.

In scalar algebra, the inverse of a number is that number which, when multiplied by the original number, gives a product of 1. Hence, the inverse of x is simple $1/x$. or, in slightly different notation, x^{-1} In matrix algebra, the inverse of a matrix is that matrix which, when multiplied by the original matrix, gives an identity matrix. Hence, $\mathbf{AA^{-1} = A^{-1}A = I}$

A matrix must be square to have an inverse, but not all square matrices have an inverse. In some cases, the inverse does not exist, that is, when the determinant equals zero (see below).

For covariance and correlation matrices, an inverse will always exist, provided that there are more subjects than there are variables and that every variable has a variance greater than 0.

Solving Systems of Equations Using Matrices

Matrices are particularly useful when solving systems of equations, which we use when we solve for the least squares estimators. Here is an example, with three equations and three unknowns:

$$x + 2y + z = 3$$
$$3x - y - 3z = -1$$
$$2x + 3y + z = 4$$

How would one go about solving this? There are various techniques, including substitution, and multiplying equations by constants and adding them to get single variables to cancel. There is an easier way, however, and that is to use a matrix. Note that this system of equations can be represented as follows:

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & -1 & -3 \\ 2 & 3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} \quad \rightarrow \quad \mathbf{Ax} = \mathbf{b}$$

We can solve the problem $\mathbf{Ax} = \mathbf{b}$ by pre-multiplying both sides by $\mathbf{A}^{-1}$ and simplifying. This yields the following:

$$\mathbf{Ax} = \mathbf{b} \rightarrow \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \quad \rightarrow \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

We can therefore solve a system of equations by computing the inverse of $\mathbf{A}$, and multiplying it by $\mathbf{b}$. Here $\mathbf{A}$ inverse is $\begin{pmatrix} 8 & 1 & -5 \\ -9 & -1 & 6 \\ 11 & 1 & -7 \end{pmatrix}$

And $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} =$ $\begin{pmatrix} 8 & 1 & -5 \\ -9 & -1 & 6 \\ 11 & 1 & -7 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 3 \end{pmatrix}$

Computationally, this is a much easier way to solve systems of equations – we just need to compute an inverse, and perform a single matrix multiplication. This approach only works, however, if the matrix **A** is nonsingular. If it is not invertable, then this will not work. In fact, if a row or a column of the matrix **A** is a linear combination of the others, there are *no* solutions to the system of equations, or *many* solutions to the system of equations. In either case, the system is said to be under-determined. We

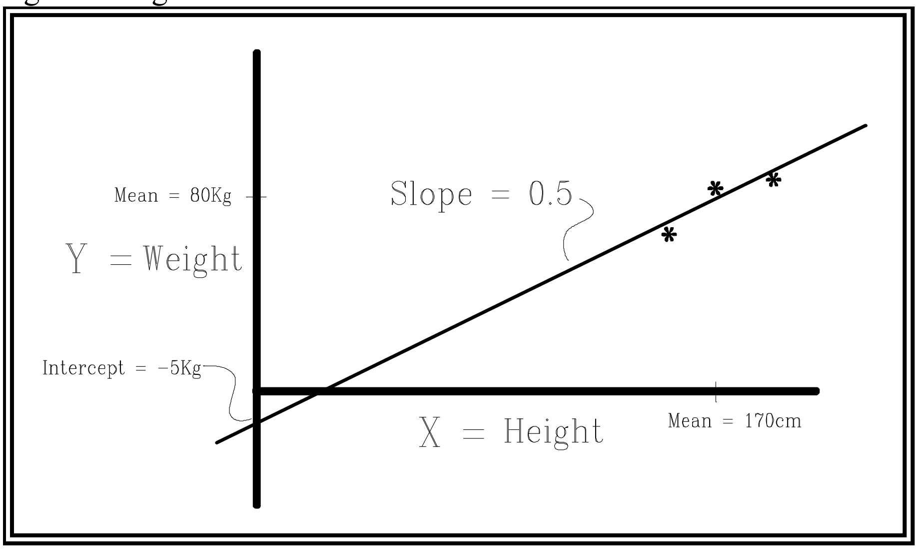
can compute the determinant of a matrix to see if it in fact is underdetermined.
Note also that for many equations, there are more efficient ways to solve such equations, using sparse matrix techniques (many coefficients in such matrices are often zero) and iteration. In act, it can be numerically quite risky to invert a very big matrix as the accumulation of very many rounding errors can become quite substantial.

Example use of Matrices: Regression.

Consider that we have a tiny data set on height and weight of individuals:

Trait	Data			Means
Weight (Y)	74	82	84	80
Height (X)	160	170	180	170

To predict weight given height we need to calculate the regression of weight on height:



$$\backslash s\backslash up5(\hat{\ }_{(\text{Y on X})} = \frac{Cov(X,\ Y)}{V_X} = \frac{\dfrac{\sum_i [\,(X_i - X)\,(Y_i - Y)\,]}{n - 1}}{\dfrac{\sum_i (X_i - X)^2}{n - 1}} =$$

$$\frac{\sum_i [\,(X_i - X)\,(Y_i - Y)\,]}{\sum_i (X_i - X)^2} = \frac{\sum_i x_i\, y_i}{\sum_i x_i^2}$$

Where x_i and y_i are height and weight expressed as deviations from their respective means.

Notice that if we make vectors $X = \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix}$ and $Y = \begin{pmatrix} -6 \\ 2 \\ 4 \end{pmatrix}$

containing deviations from

means, then notice that ...

$$X'X = \begin{pmatrix} -10 & 0 & 10 \end{pmatrix} \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix} = (200) = \sum_i x_i^2$$

$$X'Y = \begin{pmatrix} -10 & 0 & 10 \end{pmatrix} \begin{pmatrix} -6 \\ 2 \\ 4 \end{pmatrix} = (100) = \sum_i x_i y_i$$

So, just as $\hat{b} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$ we have: $\hat{b} = (X'X)^{-1} X'Y = 0.5$ in this

case.

The model we have used here is: $y_i = b \; x_i + e_i$

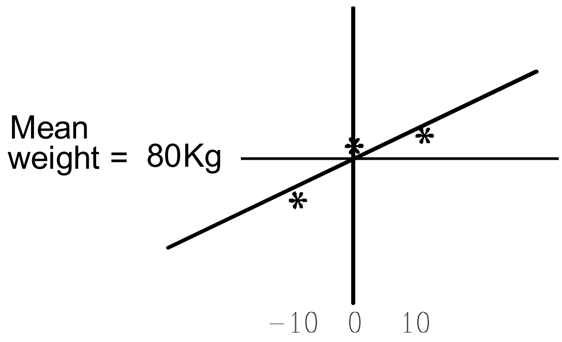
The 'e' is for error. For example @ i=1: $(-6) = 0.5\,(-10) + (-1)$
We must have an 'e' to be able to use '='.
We drop the 'e' to get predictions of weight from height: $\hat{y}_i$ = \s\up5(^ x_i $(-5) = 0.5\;(-10)$

We can write this model in matrix notation:

$$\mathbf{Y} = \mathbf{X} \quad \mathbf{b} + \mathbf{e}$$

$$\begin{pmatrix} -6 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix} \begin{pmatrix} .5 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

A more common model is:



$Y_i = b_1 \times 1 + b_2 \times (X_i -$

\s\up12(_)) + e_i

Note that the scalars Y and X are capital - not expressed as deviations from means. We now have 2 b's to be estimated:

Vector b now contains:	Matrix X now contains:
b_1 - the effect of the mean of Y (weight) The weight at height (X_i - \s\up12(_)) = 0	1 (100%) is the degree of expression of the mean's effect in *each* observation.
b_2 - the effect on Y a unit deviation in X (i.e. the regression slope of 0.5)	(X_i - \s\up12(_)) is the amount of expression of the effect of height on the i^{th} weight observation.

Now we have $\mathbf{X} = \begin{pmatrix} 1 & -10 \\ 1 & 0 \\ 1 & +10 \end{pmatrix}$ and $\mathbf{Y} = \begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix}$

The Model can be written in matrices:

$$Y = X \quad B + e$$

$$\begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix} = \begin{pmatrix} 1 & -10 \\ 1 & 0 \\ 1 & +10 \end{pmatrix} \begin{pmatrix} 80 \\ .5 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

And, as before, $\hat{b} = (X'X)^{-1} X'Y$

$$\hat{b} = (\quad X' \quad X \quad)^{-1} \quad X' \quad Y$$

\o(b\s\up5(^,\s\up5(^ $= \left(\begin{pmatrix} 1 & 1 & 1 \\ -10 & 0 & +10 \end{pmatrix} \begin{pmatrix} 1 & -10 \\ 1 & 0 \\ 1 & +10 \end{pmatrix} \right)^{-1}$

$\begin{pmatrix} 1 & 1 & 1 \\ -10 & 0 & +10 \end{pmatrix} \begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix}$

The result is \o(b\s\up5(^,\s\up5(^ $= \begin{pmatrix} 80 \\ 0.5 \end{pmatrix}$ as you would expect, mean weight 80Kg, regression slope 0.5 Kg/cm. You can check this by hand calculation, or e.g. use matrices in Excel.

A model which uses raw data is: $Y_i = b_1 \times 1 + b_2 \times X_i + e_i$

Vector b now contains:	Matrix X now contains:
b_1 - is now the intercept - the predicted value of Y (weight) at X (height) = zero	1 (100%) is the degree of expression of the intercept effect in *each* observation.
b_2 - the effect on Y a unit deviation in X (i.e. the regression slope of 0.5)	X_i is the amount of expression of the effect of height in the i^{th} observation.

Now we have $X = \begin{pmatrix} 1 & 160 \\ 1 & 170 \\ 1 & 180 \end{pmatrix}$, $Y = \begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix}$, and ...

The Model can be written in matrices:

$$Y \quad = \quad X \quad B \quad + \quad e$$

$$\begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix} = \begin{pmatrix} 1 & 160 \\ 1 & 170 \\ 1 & 180 \end{pmatrix} \begin{pmatrix} -5 \\ .5 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

Note the intercept is at -5Kg as in the original graph.

And we can predict weights (**Y**) as:

\s\up6(^ = X B

$$\begin{pmatrix} 75 \\ 80 \\ 85 \end{pmatrix} = \begin{pmatrix} 1 & 160 \\ 1 & 170 \\ 1 & 180 \end{pmatrix} \begin{pmatrix} -5 \\ .5 \end{pmatrix}$$

$$\hat{b} = (\quad X' \quad X \quad)^{-1} \quad X' \quad Y$$

\o(b \s\up5(^,\s\up5(^ =

$$\left(\begin{pmatrix} 1 & 1 & 1 \\ 160 & 170 & 180 \end{pmatrix} \begin{pmatrix} 1 & 160 \\ 1 & 170 \\ 1 & 180 \end{pmatrix} \right)^{-1}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 160 & 170 & 180 \end{pmatrix} \begin{pmatrix} 74 \\ 82 \\ 84 \end{pmatrix}$$

gives the result: \o(b \s\up5(^,\s\up5(^ = $\begin{pmatrix} -5 \\ 0.5 \end{pmatrix}$

Again this is as expected: -5Kg is the intercept - where the regression line cuts the vertical axis (this is the expected weight for x=0, i.e. for a height of zero – does this make sense-?) , and 0.5 Kg/cm is the regression slope.

> $\hat{b} = (X'X)^{-1} X'Y$ is a very powerful formula. It forms the basis of multiple regression and Analysis of Variance.
> X'X has the number of observations, X'Y has the totals
>
> With some modification, it forms the basis of BLUP.

Reference Books

Searle, S.R. 1982. Matrix Algebra Useful for Statistics.Wiley & Sons. (this books gives a llot of formal prrofs and mathematical detail)

Mrode, R.A. 1996. Linear Models for the Prediction of Animal Breeding Values. CAB Int. Oxon, UK.(The appendix is simple, similar to these notes)