PubH 7405:
REGRESSION ANALYSIS

REGRESSION IN MATRIX TERMS

A “matrix” is a display of numbers or numerical quantities laid out in a rectangular array of rows and columns. The array, or two-way table of numbers, could be rectangular or square – could be just one row (a row matrix or row vector) or one column (a column matrix or column vector). When it is square, it could be symmetric or a “diagonal matrix” (non-zero entries are on the main diagonal). The numbers of rows and of columns form the “dimension” of a matrix; for example, a “3x2” matrix has three rows and two columns.

An “entry” or “element” of a matrix need two subscripts for identification; the first for the row number and the second for the column number:

$$\mathbf{A} = [a_{ij}]$$

For example, in the following matrix we have a_{11} = 16,000 and a_{32} = 35.

$$\mathbf{A} = \begin{bmatrix} 16,000 & 23 \\ 33,000 & 47 \\ 21,000 & 35 \end{bmatrix}$$

BASIC SL REGRESSION MATRICES

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \vdots \\ 1 & x_n \end{bmatrix}$$

X is special, called "Design Matrix"

X: the "Design Matrix" for MLR:

$$\mathbf{X}_{nx(k+1)} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdot & x_{k1} \\ 1 & x_{12} & x_{22} & \cdot & x_{k2} \\ & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdot & x_{kn} \end{bmatrix}$$

First subscript: Variable, one column for each predictor; Second subscript: Subject, one row for each subject

The dimension of "Design Matrix" X is changed to handle more predictors: one column for each predictor (the number of rows is still the sample size. The first column (filled with "1") is still "optional"; not included when doing "Regression through the origin" (i.e. no intercept).

X is called the <u>Design</u> Matrix.
There are two reasons for the name:

(1) By the model, the values of X's (columns) are under the controlled of investigators: entries are fixed/designed,

(2) The design/choice is consequential: the larger the variation in x's in each column the more precise the estimate of the slope.

TRANSPOSE

The transpose of a matrix A is another matrix, denoted by A' (or A^T), that is obtained by interchanging the columns and the rows of the matrix A; that is:

$$\text{If } \mathbf{A} = \left[a_{ij}\right] \text{ then } \mathbf{A}' = \left[a_{ji}\right]$$

$$\mathbf{A}_{3x2} = \begin{bmatrix} 2 & 5 \\ 7 & 10 \\ 3 & 4 \end{bmatrix}; \mathbf{A}'_{2x3} = \begin{bmatrix} 2 & 7 & 3 \\ 5 & 10 & 4 \end{bmatrix}$$

The roles of rows and columns are <u>swapped;</u> if A is a <u>symmetric</u> matrix, $a_{ij} = a_{ji}$, we have A = A'

The transpose of a column vector is a row vector

$$\mathbf{C} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 12 \end{bmatrix}; \mathbf{C}' = \begin{bmatrix} 3 & -2 & 0 & 12 \end{bmatrix}$$

MULTIPLICATION by a Scalar

- Again, "scalar" is an (ordinary) number
- In multiplying a matrix by a scalar, every element of the matrix is multiplied by that scalar
- The result is a new matrix with the same dimension

$$\mathbf{A} = [a_{ij}]$$

$$k\mathbf{A} = [ka_{ij}]$$

MULTIPLICATION of Matrices

- Multiplication of a matrix by another matrix is much more complicated.
- First, the "order" is important; "AxB" is said to be "A is post-multiplied by B" or "B is pre-multiplied by A"; In general: AxB ≠ BxA
- There is a strong requirement on the dimensions: AxB is defined only if "the number of columns of A is equal to the number of rows of B".
- The product AxB has, as its dimension, the number of rows of A and the number of columns of B.

MULTIPLICATION FORMULA

$$\mathbf{A}_{rxc}\mathbf{B}_{cxs} = \mathbf{AB}_{rxs}$$

$$[a_{ij}]_{rxc}[b_{ij}]_{cxs} = [\sum_{\mathbf{k=1}}^{\mathbf{c}} \mathbf{a_{ik}b_{kj}}]_{rxs}$$

For entry (i,j) of AxB, we multiply row “i” of A by column “j” of B; That’s <u>why</u> the number of columns of A should be the same as the number of rows of B.

EXAMPLES

$$\begin{bmatrix} 4 & 2 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4a_1 + 2a_2 \\ 5a_1 + 8a_2 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = [2^2 + 3^2 + 5^2 = 38]$$

$$\begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 33 & 52 \\ 21 & 32 \end{bmatrix}$$

OPERATION ON SLR
BASIC DATA MATRICES

$$\mathbf{Y}'\mathbf{Y} = [y_1 \quad y_2 \quad \cdots \quad y_n]\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [\sum y_i^2]$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

"Order" is important; cannot form YX'

OPERATION ON MLR
BASIC DATA MATRICES

$$\mathbf{Y}'\mathbf{Y} = [y_1 \quad y_2 \quad \cdots \quad y_n]\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = [\sum y_i^2]$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & . & 1 \\ x_{11} & x_{12} & x_{13} & . & x_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & . & x_{kn} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \vdots \\ \sum x_{ki} y_i \end{bmatrix}$$

MORE REGRESSION EXAMPLE

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \vdots \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

Again, X is referred to as the "Design Matrix"

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \sum x_{1i} & \cdots & \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \cdots & \sum x_{1i}x_{ki} \\ \vdots & \vdots & \cdots & \vdots \\ \sum x_{ki} & \sum x_{1i}x_{ki} & \cdots & \sum x_{1i}^2 \end{bmatrix}$$

X’X is a square matrix filled with sums of squares and sums of products; we can form XX’ but it is a different n-by-n matrix which we do not need.

SIMPLE LINEAR
REGRESSION MODEL

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i ; i = 1,2,\ldots,n$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{Y}_{nx1} = \mathbf{X}_{nx2} \boldsymbol{\beta}_{2x1} + \boldsymbol{\varepsilon}_{nx1}; \text{ or}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

MULTIPLE LINEAR REGRESSION MODEL

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i ; i = 1, 2, \ldots, n$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & . & x_{k1} \\ 1 & x_{12} & x_{22} & . & x_{k2} \\ & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & . & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

MLR MODEL

IN MATRIX TERMS

$$\mathbf{Y}_{n-by-1} = \mathbf{X}_{n-by-(k+1)}\boldsymbol{\beta}_{(k+1)-by-1} + \boldsymbol{\varepsilon}_{n-by-1}$$

$$\mathbf{E}(\mathbf{Y})_{n-by-1} = \mathbf{X}\boldsymbol{\beta}$$

$$\boldsymbol{\sigma}^2(\mathbf{Y})_{n-by-n} = \sigma^2\mathbf{I}$$

OBSERVATIONS & ERRORS

$$\mathbf{Y}_{nx1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \boldsymbol{\varepsilon}_{nx1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

β: Regression Coefficient

(a column vector of parameters)

$$\boldsymbol{\beta}_{(k+1)x1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

LINEAR DEPENDENCE

Consider a set of c column vectors $C_1, C_2, \ldots, C_c$ in a rxc matrix. If we can find c scalars $k_1, k_2, \ldots, k_c$ – not all zero – so that:

$$k_1C_1 + k_2C_2 + \ldots + k_cC_c = 0$$

the c column vectors are said to be "linearly dependent". If the only set of scalars for which the above equation holds is all zero ($k_1 = \ldots = k_c = 0$), the c column vectors is linearly independent.

EXAMPLE

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 & 1 \\ 2 & 2 & 10 & 6 \\ 3 & 4 & 15 & 1 \end{bmatrix}$$

Since:

$$(5)\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + (0)\begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} + (-1)\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} + (0)\begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The four column vectors are linearly dependent

RANK OF A MATRIX

- In the previous example, the rank of $A<4$
- Since the first, second and fourth columns are linearly independent (no scalars can be found so that $k_1C_1 + k_2C_2 + k_4C_4 = 0$); the rank of that matrix A is 3.
- The **rank of a matrix** is defined as “the **maximum number of linearly independent columns** in the matrix.

SINGULAR/NON-SINGULAR MATRICES

- If the rank of a square r x r matrix A is r then matrix A is said to be nonsingular or of "full rank"
- An r x r matrix with rank less than r is said to be singular or not of full rank.

INVERSE OF A MATRIX

- The inverse of a matrix A is another matrix A^{-1} such that: $A^{-1}A = AA^{-1} = I$ where I is the identity or unit matrix.
- An inverse of a matrix is defined only for square matrices.
- Many matrices do not have an inverse; a singular matrix does not have an inverse.
- If a square matrix has an inverse, the inverse is unique; the inverse of a nonsingular or full rank matrix is also nonsingular and has same rank.

INVERSE OF A 2x2 MATRICE

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \dfrac{d}{D} & \dfrac{-b}{D} \\ \dfrac{-c}{D} & \dfrac{a}{D} \end{bmatrix}$$

$\mathrm{D} = (\mathrm{ad} - \mathrm{bc})$ is the "**determinant**" of A; or $|\mathrm{A}|$

A singular matrix does not have an inverse because its “determinants” is zero:

$$\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 7 & 21 \end{bmatrix}$$

We have :

$$(3)\begin{bmatrix} 2 \\ 7 \end{bmatrix} + (-1)\begin{bmatrix} 6 \\ 21 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and : $D = (2)(21) - (6)(7) = 0$

A SYSTEM OF EQUATIONS

It is extremely simple to write a system of equations in the matrix form – especially with many equations

$$
\begin{aligned}
3x + 4y - 10z &= 0 \\
-2x - 5y + 21z &= 14 \\
x + 29y - 2z &= -3
\end{aligned}
$$

$$
\begin{bmatrix} 3 & 4 & -10 \\ -2 & -5 & 21 \\ 1 & 29 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 14 \\ -3 \end{bmatrix}
$$

A SIMPLE APPLICATION

Consider a system of two equations :

$$2y_1 + 4y_2 = 20$$
$$3y_1 + y_2 = 10$$

which is written in matrix notation :

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Solution:

$y_1 = 2$ & $y_2 = 4$

ANOTHER EXAMPLE

$$\begin{bmatrix} 3 & 4 & -10 \\ -2 & -5 & 21 \\ 1 & 29 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 14 \\ -3 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 & 4 & -10 \\ -2 & -5 & 21 \\ 1 & 29 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 14 \\ -3 \end{bmatrix}$$

REGRESSION EXAMPLE

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

$$D = n\sum x^2 - (\sum x)(\sum x)$$

$$= n\sum (x - \bar{x})^2$$

We can easily see that (D≠0) . This property would not be true if X has more columns (Multiple Regression) and columns are not linearly independent. If “columns” (i.e. predictors/factors) are highly related, Design Matrix <u>approaching</u> “singular”: Regression failed!

REGRESSION EXAMPLE

$$\mathbf{X'X} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

$$D = n\sum(x-\bar{x})^2$$

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} \dfrac{\sum x^2}{n\sum(x-\bar{x})^2} & \dfrac{-\sum x}{n\sum(x-\bar{x})^2} \\ \dfrac{-\sum x}{n\sum(x-\bar{x})^2} & \dfrac{n}{n\sum(x-\bar{x})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum(x-\bar{x})^2} & \dfrac{-\bar{x}}{\sum(x-\bar{x})^2} \\ \dfrac{-\bar{x}}{\sum(x-\bar{x})^2} & \dfrac{1}{\sum(x-\bar{x})^2} \end{bmatrix}$$

FOUR IMPORTANT MATRICES IN REGRESSION ANALYSIS

$$\mathbf{Y}'\mathbf{Y} = [\sum y_i^2]$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix}$$

LEAST SQUARE METHOD SLR

$$\text{Data : } \{(x_i, y_i)\}_{i=1}^{n}$$

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\delta Q}{\delta \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\delta Q}{\delta \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

LEAST SQUARE METHOD MLR

$$\text{Data}: \{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)\}_{i=1}^{n}$$

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki})^2$$

We solve equations :

$$\frac{\delta Q}{\delta \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = 0$$

$$\frac{\delta Q}{\delta \beta_1} = -2\sum_{i=1}^{n} x_{1i}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = 0$$

$$\ldots$$

$$\frac{\delta Q}{\delta \beta_k} = -2\sum_{i=1}^{n} x_{ki}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = 0$$

NORMAL EQUATIONS

$$\frac{\delta Q}{\delta \beta_0} = \frac{\delta Q}{\delta \beta_1} = 0$$

Normal Equations :

$$\sum \mathrm{y_i} = nb_0 + b_1 \sum x_i$$

$$\sum \mathrm{x_i y_i} = b_0 \sum x_i + b_1 \sum x_i^2$$

In matrix notations :

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\mathbf{X'X}_{2x2}\mathbf{b}_{2x1} = \mathbf{X'Y}_{2x1}$$

NORMAL EQUATIONS IN MLR

$$\frac{\delta Q}{\delta \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = \sum_{i=1}^{n} e_i = 0$$

$$\frac{\delta Q}{\delta \beta_1} = -2\sum_{i=1}^{n} x_{1i}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = \sum_{i=1}^{n} x_{1i} e_i = 0$$

$$\ldots$$

$$\frac{\delta Q}{\delta \beta_k} = -2\sum_{i=1}^{n} x_{ik}(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} \ldots - \beta_k x_{ki}) = \sum_{i=1}^{n} x_{ki} e_i = 0$$

In MLR, these normal equations look more complicated but will lead to the same result in matrix terms:

$$(X'X)b = X'Y$$

LEAST SQUARE ESTIMATES

Normal Equations :

$$\sum \mathrm{y_i} = nb_0 + b_1 \sum x_i$$

$$\sum \mathrm{x_i y_i} = b_0 \sum x_i + b_1 \sum x_i^2$$

In matrix notations :

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$$

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

SUMMARY REGRESSION RESULTS

Normal Equations :

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$$

Least Square Estimates :

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

WE PROVE THE SAME RESULTS

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix}$$

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$b_1 = \frac{\sum x_i y_i - \left[\sum x_i\right]\left[\sum y_i\right]/n}{\sum (x-\bar{x})^2}$$

MORE DIRECT APPROACH

Instead of normal equations, we could start earlier with the Sum of Squared Errors (SSE)

Sum of squared errors :

$$Q=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{1i}-\beta_2x_{2i}\ldots-\beta_kx_{ki})^2$$

In matrix notation :

$$Q=(\mathbf{Y}-\mathbf{X\beta})^{'}(\mathbf{Y}-\mathbf{X\beta})$$

$$=\mathbf{Y}^{'}\mathbf{Y}-2\mathbf{\beta}^{'}\mathbf{X}^{'}\mathbf{Y}+\mathbf{\beta}^{'}\mathbf{X}^{'}\mathbf{X\beta}$$

Normal Equations :

$$Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{'}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \mathbf{Y}^{'}\mathbf{Y} - 2\boldsymbol{\beta}^{'}\mathbf{X}^{'}\mathbf{Y} + \boldsymbol{\beta}^{'}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}$$

$$\frac{\boldsymbol{\delta}}{\boldsymbol{\delta\beta}}(Q) = \begin{bmatrix} \dfrac{\delta Q}{\delta \beta_0} \\ \dfrac{\delta Q}{\delta \beta_1} \\ \vdots \\ \dfrac{\delta Q}{\delta \beta_k} \end{bmatrix}$$

$$= -2\mathbf{X}^{'}\mathbf{Y} + 2\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}$$

$$= 0 \Leftrightarrow \mathbf{X}^{'}\mathbf{X}\mathbf{b} = \mathbf{X}^{'}\mathbf{Y}$$

We can prove the "normal equations" by taking derivative of SSE:

$SSE = Y'Y - 2\beta'X'Y + \beta'X'X\beta$

To do that, we need to learn:

(1) Derivative of $\beta'[X'Y]$ is $X'Y$

(2) Derivative of $\beta'[X'X]\beta$ is $2[X'X]\beta$

Putting together :

$$Q = (\mathbf{Y} - \mathbf{X\beta})^{'}(\mathbf{Y} - \mathbf{X\beta})$$

$$= (\mathbf{Y' - \beta' X'})(\mathbf{Y} - \mathbf{X\beta})$$

$$= \mathbf{Y}^{'}\mathbf{Y} - 2\boldsymbol{\beta}^{'}\mathbf{X}^{'}\mathbf{Y} + \boldsymbol{\beta}^{'}\mathbf{X}^{'}\mathbf{X\beta}$$

$$\frac{\boldsymbol{\delta}}{\boldsymbol{\delta\beta}}(Q) = \begin{bmatrix} \dfrac{\delta Q}{\delta \beta_0} \\ \dfrac{\delta Q}{\delta \beta_1} \end{bmatrix}$$

$$= -2\mathbf{X}^{'}\mathbf{Y} + 2\mathbf{X}^{'}\mathbf{X\beta}$$

Note :

$$= 0 \Leftrightarrow (\mathbf{X}^{'}\mathbf{X})\mathbf{b} = \mathbf{X'Y}$$

LEAST SQUARE ESTIMATE

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$$

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

Note the error in equation (6.25) of the text

THE HAT MATRIX

$$\hat{\mathbf{Y}} = \mathbf{Xb}$$

$$= \mathbf{X}[(\mathbf{X'X})^{-1}\mathbf{X'Y}]$$

$$= [\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]\mathbf{Y}$$

$$= \mathbf{HY}$$

$\mathbf{H} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$ is called the "Hat Matrix"

IDEMPOTENCY

the "Hat Matrix" :

$$\mathbf{H} = \mathbf{X}(\mathbf{X'}\,\mathbf{X})^{-1}\,\mathbf{X'}$$

is "idempotent":

$$\mathbf{HH} = \mathbf{H}$$

RANDOM VECTORS & MATRICES

A random vector or a random matrix contains elements which are random variables.

RANDOM VECTORS IN SLR

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

EXPECTED VALUES

$$\mathbf{Y} = \mathbf{E(Y)} + \boldsymbol{\varepsilon}$$

$$\mathbf{E(Y)} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \mathrm{M} \\ E(Y_n) \end{bmatrix}; \mathbf{E(\boldsymbol{\varepsilon})} = \begin{bmatrix} 0 \\ 0 \\ \mathrm{M} \\ 0 \end{bmatrix}$$

VARIANCE-COVARIANCE MATRIX

The variances (of elements of a random matrix) and the covariance between any two elements (of elements of a random matrix) are assembled in the variance-covariance matrix – denoted by either Var(Y) or $\sigma^2(Y)$ – or ∑2

EXAMPLE: (BIVARIATE) VECTOR

$$\text{Variable} : \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$\text{Mean} : \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\text{Variance} - \text{Covariance Matrix} : \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

We have for random <u>variables</u>:

$E(a_* Y) = a_* E(Y)$

$Var(a_* Y) = a^2{}_* var(Y)$

What about random <u>vectors</u>? Say:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$
\begin{aligned}
\mathbf{AY} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \\
&= \begin{bmatrix} a_{11}Y_1 + a_{12}Y_2 \\ a_{21}Y_1 + a_{22}Y_2 \end{bmatrix} \\
E(\mathbf{AY}) &= \begin{bmatrix} E(a_{11}Y_1 + a_{12}Y_2) \\ E(a_{21}Y_1 + a_{22}Y_2) \end{bmatrix} \\
&= \begin{bmatrix} a_{11}E(Y_1) + a_{12}E(Y_2) \\ a_{21}E(Y)_1 + a_{22}E(Y_2) \end{bmatrix} \\
&= \mathbf{AE(Y)}
\end{aligned}
$$

$$\mathbf{AY} = \begin{bmatrix} a_{11}Y_1 + a_{12}Y_2 \\ a_{21}Y_1 + a_{22}Y_2 \end{bmatrix}$$

$$\sigma^2(\mathbf{AY}) = \begin{bmatrix} \sigma^2(a_{11}Y_1 + a_{12}Y) & \sigma(a_{11}Y_1 + a_{12}Y, a_{21}Y_1 + a_{22}Y) \\ \sigma(a_{11}Y_1 + a_{12}Y, a_{21}Y_1 + a_{22}Y) & \sigma^2(a_{21}Y_1 + a_{22}Y) \end{bmatrix}$$

$$= \text{L}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) \\ \sigma(Y_1, Y_2) & \sigma^2(Y_2) \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$$= \mathbf{A}\boldsymbol{\sigma}^2(\mathbf{Y})\mathbf{A}'$$

Can verify backward too, starting with Aσ^2(Y)A'

REGRESSION EXAMPLES

$$\mathbf{Y} = \mathbf{E(Y)} + \boldsymbol{\varepsilon}$$

$$\mathbf{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

$$\mathbf{Var(Y)} = \begin{bmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \cdots & \sigma(Y_1, Y_n) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \cdots & \sigma(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma(Y_n, Y_1) & \sigma(Y_n, Y_2) & \cdots & \sigma^2(Y_n) \end{bmatrix}$$

$$= \sigma^2\mathbf{I}$$

FITTED VALUES

$$\hat{\mathbf{Y}} = \begin{bmatrix} b_0 + b_1 x_1 \\ b_0 + b_1 x \\ \vdots \\ b_0 + b_1 x_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{Xb}$$

RESIDUALS

Model:

$$\mathbf{Y}_{nx1} = \mathbf{X}_{nx2}\boldsymbol{\beta}_{2x1} + \boldsymbol{\varepsilon}_{nx1}$$

Fitted Value:

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$$

Residuals:

$$\begin{aligned}\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\ &= \mathbf{Y} - \mathbf{X}\mathbf{b} \\ &= \mathbf{Y} - \mathbf{H}\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{Y}\end{aligned}$$

Like the Hat Matrix H, (I-H) is symmetric & idempotent

VARIANCE OF RESIDUALS

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\boldsymbol{\sigma}^2(\mathbf{e}) = (\mathbf{I} - \mathbf{H})\boldsymbol{\sigma}^2(\mathbf{Y})(\mathbf{I} - \mathbf{H})'$$

$$= (\mathbf{I} - \mathbf{H})(\sigma^2\mathbf{I})(\mathbf{I} - \mathbf{H})'$$

$$= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})'$$

$$= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})$$

$$= \sigma^2(\mathbf{I} - \mathbf{H})$$

$$\hat{=} MSE(\mathbf{I} - \mathbf{H})$$

REGRESSION COEFFICIENTS

$$\frac{\boldsymbol{\delta}}{\boldsymbol{\delta\beta}}(Q) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$= 0 \Leftrightarrow \mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\mathbf{b}$$

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

$$= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

$$= \mathbf{A}\mathbf{Y}$$

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

VARIANCE OF REGRESSION COEFFICIENTS

$$\begin{aligned}
\mathbf{b} &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} \\
&= \mathbf{A}\mathbf{Y} \\
\boldsymbol{\sigma}^2(\mathbf{b}) &= \mathbf{A}\boldsymbol{\sigma}^2(\mathbf{Y})\mathbf{A}' \\
&= \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}' \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\
\mathbf{s}^2(\mathbf{b}) &= MSE(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

$$\mathbf{b} = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

$$= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

$$= \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum(x-\bar{x})^2} & \frac{-\bar{x}}{\sum(x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum(x-\bar{x})^2} & \frac{1}{\sum(x-\bar{x})^2} \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Can verify : same results, eg.

$$\mathrm{b}_1 = \frac{-\bar{x}\sum y + \sum xy}{\sum(x-\bar{x})^2}$$

$$= \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

<u>EXAMPLE</u>: SL REGRESSION

$$\mathbf{s}^2(\mathbf{b}) = MSE(\mathbf{X}'\mathbf{X})^{-1}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix}$$

$$\mathbf{s}^2(\mathbf{b}) = MSE \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix}$$

$$\mathbf{s}^2(\mathbf{b}) = MSE \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2} & \frac{-\bar{x}}{\sum (x-\bar{x})^2} \\ \frac{-\bar{x}}{\sum (x-\bar{x})^2} & \frac{1}{\sum (x-\bar{x})^2} \end{bmatrix}$$

$$s^2(b_0) = MSE\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x-\bar{x})^2}\right]$$

$$s^2(b_1) = \frac{MSE}{\sum (x-\bar{x})^2}$$

$$s(b_0, b_1) = \frac{-MSE(\bar{x})}{\sum (x-\bar{x})^2}$$

Same results for Variances;

Covariance is new - Only need Mean & Variance of X

THE MEAN RESPONSE

Let $X = x_h$ denote the level of X for which we wish to estimate the mean response, i.e. $E(Y|X=x_h)$. The only thing new is that X and x_h are vector; $x_h = (x_{1h}, x_{2h},\ldots, x_{kh})$. The point estimate of the response is:

$$E(Y \mid \mathbf{X} = \mathbf{x}_h) = \hat{Y}_h$$
$$= b_0 + b_1 x_{1h} + b_2 x_{2h} + \ldots + b_k x_{kh}$$

In matrix terms:

$$\hat{Y} = \mathbf{X_h'b}$$

$$\sigma^2(\hat{Y}) = \mathbf{X_h'}\boldsymbol{\sigma}^2(\mathbf{b})\mathbf{X_h}$$

$$= \sigma^2 \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h}$$

$$s^2(\hat{Y}) = MSE(\mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h})$$

PREDICTION OF NEW OBSERVATION

Let X = x_h denote the level of X under investigation, at which the mean response is $E(Y|X=x_h)$. Let $Y_{h(new)}$ be the value of the new individual response of interest. The point estimate is still the same $E(Y|X=x_h)$:

$$\hat{Y}_{h(new)} = b_0 + b_1 x_{1h} + b_2 x_{2h} + \ldots + b_k x_{kh}$$
$$= \mathbf{X}_{\mathbf{h}}^{'}\mathbf{b}$$

$$\mathbf{Var}(\hat{Y}_{h(new)}) = \sigma^2\{1 + \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h}\}$$

$$\mathbf{s}^2(\hat{Y}_{h(new)}) = MSE\{1 + \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h}\}$$

The matrix notation and derivations may be deceiving because <u>they hide enormous computational complexities</u>. To find the inverse of a 10x10 matrix, for example X'X with k = 9, requires tremendous amount of computation. However, the actual computations will be done by computer; hence it does not matter to us whether X'X represents a 2x2 or a 6x6 matrix.

Example:

SBP versus AGE

	Age (x)	SBP (y)	x-sq	y-sq	xy
	42	130	1764	16900	5460
	46	115	2116	13225	5290
	42	148	1764	21904	6216
	71	100	5041	10000	7100
	80	156	6400	24336	12480
	74	162	5476	26244	11988
	70	151	4900	22801	10570
	80	156	6400	24336	12480
	85	162	7225	26244	13770
	72	158	5184	24964	11376
	64	155	4096	24025	9920
	81	160	6561	25600	12960
	41	125	1681	15625	5125
	61	150	3721	22500	9150
	75	165	5625	27225	12375
Totals	984	2193	67954	325929	146260

$$\mathbf{X'X} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} = \begin{bmatrix} 15 & 984 \\ 984 & 67{,}954 \end{bmatrix}$$

$$D = (15)(67{,}954) - (984)^2 = 51{,}054$$

$$\mathbf{X'X}^{-1} = \begin{bmatrix} \dfrac{67{,}954}{51{,}064} & \dfrac{-984}{51{,}064} \\ \dfrac{-984}{51{,}064} & \dfrac{15}{51{,}064} \end{bmatrix}$$

$$= \begin{Bmatrix} 1.3310 & -.0193 \\ -.0193 & .0003 \end{Bmatrix}$$

$$\mathbf{X'X} = \begin{bmatrix} 15 & 984 \\ 984 & 67{,}954 \end{bmatrix}$$

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 1.3310 & -.0193 \\ -.0193 & .0003 \end{bmatrix}$$

$$\mathbf{X'Y} = \begin{bmatrix} 2{,}193 \\ 146{,}260 \end{bmatrix}$$

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

$$= \begin{bmatrix} 1.3310 & -.0193 \\ -.0193 & .0003 \end{bmatrix} \begin{bmatrix} 2{,}193 \\ 146{,}260 \end{bmatrix}$$

VARIANCE OF REGRESSION COEFFICIENTS

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

$$\mathbf{s}^2(\mathbf{b}) = MSE(\mathbf{X}'\mathbf{X})^{-1}$$

$$= MSE\begin{bmatrix} 1.3310 & -.0193 \\ -.0193 & .0003 \end{bmatrix}$$

SUMMARIES

- **All results can be put in the matrix forms**
- **If we can inverse a matrix and can multiply two matrices, we can get all numerical results – even without a packaged computer program.**
- **In matrix their forms, results can be easier generalized; the only change needed is the Design Matrix (& its dimension) so as to handle more than one predictors.**

Readings & Exercises

- Readings: A thorough reading of the text's sections 5.1-5.13 (pp.176-209) and sections 6.2-6.9 (pp.222-247) is highly recommended.
- Exercises: The following exercises are good for practice, all from chapter 5 of text: 5.1,5.2, 5.7-5.11 and 5.24-5.26; plus these from chapter 6 of text: 6.5(b-d), 6.7, 6.10(a-d), and 6.15(a-f).

Due As Homework

x	y
24	38.8
28	39.5
32	40.3
36	40.7
40	41.0
44	41.1
48	41.4
52	41.6
56	41.8
60	41.9

#12.1 The following data were collected during an experiment in which 10 laboratory animals were inoculated with a pathogen. The variables are Time after inoculation (X, in minutes) and Temperature (Y, in Celsius degrees).

For the regression of Y (as dependent variable) on X (as sole predictor), form these matrices:

Y'Y, X'Y, and X'X

#12.2 Solve the following system of equations:

$$\mathbf{7x - 6y = 12}$$

$$\mathbf{3x + 9y = 25}$$