Introduction to Matrices and Matrix Approach to Simple Linear Regression

Matrices

- Definition: A matrix is a rectangular array of numbers or symbolic elements
- In many applications, the rows of a matrix will represent individuals cases (people, items, plants, animals,...) and columns will represent attributes or characteristics
- The dimension of a matrix is it number of rows and columns, often denoted as *r* x *c* (*r* rows by *c* columns)
- Can be represented in full form or abbreviated form:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix} = \begin{bmatrix} a_{ij} \end{bmatrix} \quad i=1,\ldots,r;\ j=1,\ldots,c$$

Special Types of Matrices

Square Matrix: Number of rows = # of Columns $(r = c)$

$$\mathbf{A} = \begin{bmatrix} 20 & 32 & 50 \\ 12 & 28 & 42 \\ 28 & 46 & 60 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Vector: Matrix with one column (column vector) or one row (row vector)

$$\mathbf{C} = \begin{bmatrix} 57 \\ 24 \\ 18 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \qquad \mathbf{E}' = \begin{bmatrix} 17 & 31 \end{bmatrix} \qquad \mathbf{F}' = \begin{bmatrix} f_1 & f_2 & f_3 \end{bmatrix}$$

Transpose: Matrix formed by interchanging rows and columns of a matrix (use "prime" to denote transpose)

$$\underset{2\times 3}{\mathbf{G}} = \begin{bmatrix} 6 & 15 & 22 \\ 8 & 13 & 25 \end{bmatrix} \qquad \underset{3\times 2}{\mathbf{G}'} = \begin{bmatrix} 6 & 8 \\ 15 & 13 \\ 22 & 25 \end{bmatrix}$$

$$\underset{r\times c}{\mathbf{H}} = \begin{bmatrix} h_{11} & \cdots & h_{1c} \\ \vdots & & \vdots \\ h_{r1} & \cdots & h_{rc} \end{bmatrix} = \begin{bmatrix} h_{ij} \end{bmatrix} \quad i = 1,\dots,r;\ j = 1,\dots,c \quad \Rightarrow \quad \underset{c\times r}{\mathbf{H}'} = \begin{bmatrix} h_{11} & \cdots & h_{r1} \\ \vdots & & \vdots \\ h_{1c} & \cdots & h_{rc} \end{bmatrix} = \begin{bmatrix} h_{ji} \end{bmatrix} \quad j = 1,\dots,c;\ i = 1,\dots,r$$

Matrix Equality: Matrices of the same dimension, and corresponding elements in same cells are all equal:

$$\mathbf{A} = \begin{bmatrix} 4 & 6 \\ 12 & 10 \end{bmatrix} = \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \Rightarrow \quad b_{11} = 4,\ b_{12} = 6,\ b_{21} = 12,\ b_{22} = 10$$

Regression Examples - Toluca Data

Response Vector: $\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$

$$\underset{1\times n}{\mathbf{Y}'} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix}$$

Design Matrix: $\underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$

$$\underset{2\times n}{\mathbf{X}'} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}$$

X			Y
1	80		399
1	30		121
1	50		221
1	90		376
1	70		361
1	60		224
1	120		546
1	80		352
1	100		353
1	50		157
1	40		160
1	70		252
1	90		389
1	20		113
1	110		435
1	100		420
1	30		212
1	50		268
1	90		377
1	110		421
1	30		273
1	90		468
1	40		244
1	80		342
1	70		323

Matrix Addition and Subtraction

Addition and Subtraction of 2 Matrices of Common Dimension:

$$\underset{2\times 2}{\mathbf{C}} = \begin{bmatrix} 4 & 7 \\ 10 & 12 \end{bmatrix} \qquad \underset{2\times 2}{\mathbf{D}} = \begin{bmatrix} 2 & 0 \\ 14 & 6 \end{bmatrix} \qquad \underset{2\times 2}{\mathbf{C}+\mathbf{D}} = \begin{bmatrix} 4+2 & 7+0 \\ 10+14 & 12+6 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 24 & 18 \end{bmatrix} \qquad \underset{2\times 2}{\mathbf{C}-\mathbf{D}} = \begin{bmatrix} 4-2 & 7-0 \\ 10-14 & 12-6 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ -4 & 6 \end{bmatrix}$$

$$\underset{r\times c}{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix} = \begin{bmatrix} a_{ij} \end{bmatrix} \quad i=1,\dots,r;\ j=1,\dots,c \qquad \underset{r\times c}{\mathbf{B}} = \begin{bmatrix} b_{11} & \cdots & b_{1c} \\ \vdots & & \vdots \\ b_{r1} & \cdots & b_{rc} \end{bmatrix} = \begin{bmatrix} b_{ij} \end{bmatrix} \quad i=1,\dots,r;\ j=1,\dots,c$$

$$\underset{r\times c}{\mathbf{A}+\mathbf{B}} = \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1c}+b_{1c} \\ \vdots & & \vdots \\ a_{r1}+b_{r1} & \cdots & a_{rc}+b_{rc} \end{bmatrix} = \begin{bmatrix} a_{ij}+b_{ij} \end{bmatrix} \quad i=1,\dots,r;\ j=1,\dots,c$$

$$\underset{r\times c}{\mathbf{A}-\mathbf{B}} = \begin{bmatrix} a_{11}-b_{11} & \cdots & a_{1c}-b_{1c} \\ \vdots & & \vdots \\ a_{r1}-b_{r1} & \cdots & a_{rc}-b_{rc} \end{bmatrix} = \begin{bmatrix} a_{ij}-b_{ij} \end{bmatrix} \quad i=1,\dots,r;\ j=1,\dots,c$$

Regression Example:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \underset{n\times 1}{\mathbf{Y}} = \underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \quad \text{since} \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\}+\varepsilon_1 \\ E\{Y_2\}+\varepsilon_2 \\ \vdots \\ E\{Y_n\}+\varepsilon_n \end{bmatrix}$$

Matrix Multiplication

Multiplication of a Matrix by a Scalar (single number):

$$k = 3 \quad \mathbf{A} = \begin{bmatrix} 2 & 1 \\ -2 & 7 \end{bmatrix} \quad \Rightarrow \quad k\mathbf{A} = \begin{bmatrix} 3(2) & 3(1) \\ 3(-2) & 3(7) \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ -6 & 21 \end{bmatrix}$$

Multiplication of a Matrix by a Matrix (#cols(**A**) = #rows(**B**)):

$$\text{If } \; c_A = r_B : \underset{r_A \times c_A}{\mathbf{A}} \; \underset{r_B \times c_B}{\mathbf{B}} = \underset{r_A \times c_B}{\mathbf{AB}} = \left[ab_{ij} \right] \; i = 1,...,r_A; \; j = 1,...,c_B$$

$ab_{ij} \equiv$ sum of the products of the $c_A = r_B$ elements of i^th^ row of **A** and j^th^ column of **B**:

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 3 & -1 \\ 0 & 7 \end{bmatrix} \quad \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 3 & -1 \\ 2 & 4 \end{bmatrix}$$

$$\underset{3\times 2}{\mathbf{A}} \; \underset{2\times 2}{\mathbf{B}} = \underset{3\times 2}{\mathbf{AB}} = \begin{bmatrix} 2(3)+5(2) & 2(-1)+5(4) \\ 3(3)+(-1)(2) & 3(-1)+(-1)(4) \\ 0(3)+7(2) & 0(-1)+7(4) \end{bmatrix} = \begin{bmatrix} 16 & 18 \\ 7 & -7 \\ 14 & 8 \end{bmatrix}$$

$$\text{If } \; c_A = r_B = c : \underset{r_A \times c_A}{\mathbf{A}} \; \underset{r_B \times c_B}{\mathbf{B}} = \underset{r_A \times c_B}{\mathbf{AB}} = \left[ab_{ij} \right] = \left[\sum_{k=1}^{c} a_{ik} b_{kj} \right] \; i = 1,...,r_A; \; j = 1,...,c_B$$

Matrix Multiplication Examples - I

Simultaneous Equations: $a_{11}x_1 + a_{12}x_2 = y_1 \quad a_{21}x_1 + a_{22}x_2 = y_2$

(2 equations: x_1, x_2 unknown): $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$$\Rightarrow \quad \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{AX} = \mathbf{Y}$$

Sum of Squares: $4^2 + \left(-2\right)^2 + 3^2 = \begin{bmatrix} 4 & -2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 29 \end{bmatrix}$

Regression Equation (Expected Values): $\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$

Matrix Multiplication Examples - II

Matrices used in simple linear regression (that generalize to multiple regression):

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{i=1}^{n} Y_i^2$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix} \qquad \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

Special Matrix Types

Symmetric Matrix: Square matrix with a transpose equal to itself: **A = A'**:

$$\mathbf{A} = \begin{bmatrix} 6 & 19 & -8 \\ 19 & 14 & 3 \\ -8 & 3 & 1 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 6 & 19 & -8 \\ 19 & 14 & 3 \\ -8 & 3 & 1 \end{bmatrix} = \mathbf{A}$$

Diagonal Matrix: Square matrix with all off-diagonal elements equal to 0:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} \qquad \text{Note: Diagonal matrices are symmetric (not vice versa)}$$

Identity Matrix: Diagonal matrix with all diagonal elements equal to 1 (acts like multiplying a scalar by 1):

$$\underset{3\times 3}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow \mathbf{IA} = \mathbf{AI} = \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Scalar Matrix: Diagonal matrix with all diagonal elements equal to a single number"

$$\begin{bmatrix} k & 0 & 0 & 0 \\ 0 & k & 0 & 0 \\ 0 & 0 & k & 0 \\ 0 & 0 & 0 & k \end{bmatrix} = k \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = k \underset{4\times 4}{\mathbf{I}}$$

1-Vector and matrix and zero-vector:

$$\underset{r\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \underset{r\times r}{\mathbf{J}} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \qquad \underset{r\times 1}{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\text{Note: } \underset{1\times r}{\mathbf{1}'}\,\underset{r\times 1}{\mathbf{1}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = r \qquad \underset{r\times 1}{\mathbf{1}}\,\underset{1\times r}{\mathbf{1}'} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \underset{r\times r}{\mathbf{J}}$$

Linear Dependence and Rank of a Matrix

- Linear Dependence: When a linear function of the columns (rows) of a matrix produces a zero vector (one or more columns (rows) can be written as linear function of the other columns (rows))
- Rank of a matrix: Number of linearly independent columns (rows) of the matrix. Rank cannot exceed the minimum of the number of rows or columns of the matrix. rank(**A**) ≤ min(r_A,c_a)
- A matrix if full rank if rank(**A**) = min(r_A,c_a)

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} 1 & -3 \\ -4 & 12 \end{bmatrix} = \begin{bmatrix} \underset{2\times 1}{\mathbf{A}_1} & \underset{2\times 1}{\mathbf{A}_2} \end{bmatrix} \qquad 3\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{0} \quad \text{Columns of } \mathbf{A} \text{ are linearly dependent} \quad \text{rank}(\mathbf{A}) = 1$$

$$\underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 4 & -3 \\ 4 & 12 \end{bmatrix} = \begin{bmatrix} \underset{2\times 1}{\mathbf{B}_1} & \underset{2\times 1}{\mathbf{B}_2} \end{bmatrix} \qquad 0\mathbf{B}_1 + 0\mathbf{B}_2 = \mathbf{0} \quad \text{Columns of } \mathbf{B} \text{ are linearly independent} \quad \text{rank}(\mathbf{B}) = 2$$

Matrix Inverse

- Note: For scalars (except 0), when we multiply a number, by its reciprocal, we get 1:
  $2(1/2)=1 \qquad x(1/x)=x(x^{-1})=1$
- In matrix form if **A** is a square matrix and full rank (all rows and columns are linearly independent), then **A** has an inverse: $\mathbf{A}^{-1}$ such that: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

$$\mathbf{A} = \begin{bmatrix} 2 & 8 \\ 4 & -2 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} \frac{2}{36} & \frac{8}{36} \\ \frac{4}{36} & \frac{-2}{36} \end{bmatrix} \qquad \mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} \frac{2}{36} & \frac{8}{36} \\ \frac{4}{36} & \frac{-2}{36} \end{bmatrix} \begin{bmatrix} 2 & 8 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} \frac{4}{36}+\frac{32}{36} & \frac{16}{36}-\frac{16}{36} \\ \frac{8}{36}-\frac{8}{36} & \frac{32}{36}+\frac{4}{36} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

$$\mathbf{B} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix} \quad \mathbf{B}^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & -1/2 & 0 \\ 0 & 0 & 1/6 \end{bmatrix} \quad \mathbf{B}\mathbf{B}^{-1} = \begin{bmatrix} 4(1/4)+0+0 & 0+0+0 & 0+0+0 \\ 0+0+0 & 0+(-2)(-1/2)+0 & 0+0+0 \\ 0+0+0 & 0+0+0 & 0+0+6(1/6) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}$$

Computing an Inverse of 2x2 Matrix

$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \equiv$ full rank (columns/rows are linearly independent)

Determinant of $\mathbf{A} \equiv |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$

Note: If A is not full rank (for some value k): $a_{11} = ka_{12} \quad a_{21} = ka_{22}$

$\Rightarrow |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21} = ka_{12}a_{22} - a_{12}ka_{22} = 0$

$\underset{2\times 2}{\mathbf{A}^{-1}} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$ Thus $\mathbf{A}^{-1}$ does not exist if $\mathbf{A}$ is not full rank

While there are rules for general $r \times r$ matrices, we will use computers to solve them

Regression Example:

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \Rightarrow \mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix} \Rightarrow |\mathbf{X'X}| = n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2 = n\left(\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}\right) = n\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$$

$$\Rightarrow (\mathbf{X'X})^{-1} = \frac{1}{n\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}\begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix}$$

Note: $\sum_{i=1}^n X_i = n\overline{X} \qquad \sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2 \Rightarrow \sum_{i=1}^n X_i^2 = \sum_{i=1}^n \left(X_i - \overline{X}\right)^2 + n\overline{X}^2$

$$\Rightarrow (\mathbf{X'X})^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} & -\frac{\overline{X}}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} \\ -\frac{\overline{X}}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} & \frac{1}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} \end{bmatrix}$$

Use of Inverse Matrix – Solving Simultaneous Equations

AY = C where **A** and **C** are matrices of of constants, **Y** is matrix of unknowns

$\Rightarrow \mathbf{A}^{-1}\mathbf{A}\mathbf{Y} = \mathbf{A}^{-1}\mathbf{C} \quad \Rightarrow \quad \mathbf{Y} = \mathbf{A}^{-1}\mathbf{C}$ (assuming **A** is square and full rank)

Equation 1: $12y_1 + 6y_2 = 48$ Equation 2: $10y_1 - 2y_2 = 12$

$$\mathbf{A} = \begin{bmatrix} 12 & 6 \\ 10 & -2 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 48 \\ 12 \end{bmatrix} \qquad \mathbf{Y} = \mathbf{A}^{-1}\mathbf{C}$$

$$\Rightarrow \quad \mathbf{A}^{-1} = \frac{1}{12(-2) - 6(10)} \begin{bmatrix} -2 & -6 \\ -10 & 12 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 2 & 6 \\ 10 & -12 \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{A}^{-1}\mathbf{C} = \frac{1}{84} \begin{bmatrix} 2 & 6 \\ 10 & -12 \end{bmatrix} \begin{bmatrix} 48 \\ 12 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 96 + 72 \\ 480 - 144 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 168 \\ 336 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Note the wisdom of waiting to divide by $|\mathbf{A}|$ at end of calculation!

Useful Matrix Results

All rules assume that the matrices are conformable to operations:

Addition Rules:

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

Multiplication Rules:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \qquad \mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \qquad k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B} \quad k \equiv \text{ scalar}$$

Transpose Rules:

$$(\mathbf{A}')' = \mathbf{A} \qquad (\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \qquad (\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \qquad (\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$$

Inverse Rules (Full Rank, Square Matrices):

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \qquad (\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \qquad (\mathbf{A}^{-1})^{-1} = \mathbf{A} \qquad (\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

Random Vectors and Matrices

Shown for case of n=3, generalizes to any n:

Random variables: $Y_1, Y_2, Y_3 \quad \Rightarrow \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$

Expectation: $\mathbf{E}\left\{\mathbf{Y}\right\} = \begin{bmatrix} E\left\{Y_1\right\} \\ E\left\{Y_2\right\} \\ E\left\{Y_3\right\} \end{bmatrix}$ In general: $\underset{n\times p}{\mathbf{E}\left\{\mathbf{Y}\right\}} = \left[E\left\{Y_{ij}\right\}\right] \quad i=1,...,n;\ j=1,...,p$

Variance-Covariance Matrix for a Random Vector:

$$\sigma^2\left\{\mathbf{Y}\right\} = E\left\{\left[\mathbf{Y}-\mathbf{E}\left\{\mathbf{Y}\right\}\right]\left[\mathbf{Y}-\mathbf{E}\left\{\mathbf{Y}\right\}\right]'\right\} = \mathbf{E}\left\{\begin{bmatrix} Y_1 - E\left\{Y_1\right\} \\ Y_2 - E\left\{Y_2\right\} \\ Y_3 - E\left\{Y_3\right\} \end{bmatrix}\begin{bmatrix} Y_1 - E\left\{Y_1\right\} & Y_2 - E\left\{Y_2\right\} & Y_3 - E\left\{Y_3\right\} \end{bmatrix}\right\} =$$

$$= \mathbf{E}\left\{\begin{bmatrix} \left(Y_1 - E\left\{Y_1\right\}\right)^2 & \left(Y_1 - E\left\{Y_1\right\}\right)\left(Y_2 - E\left\{Y_2\right\}\right) & \left(Y_1 - E\left\{Y_1\right\}\right)\left(Y_3 - E\left\{Y_3\right\}\right) \\ \left(Y_2 - E\left\{Y_2\right\}\right)\left(Y_1 - E\left\{Y_1\right\}\right) & \left(Y_2 - E\left\{Y_2\right\}\right)^2 & \left(Y_2 - E\left\{Y_2\right\}\right)\left(Y_3 - E\left\{Y_3\right\}\right) \\ \left(Y_3 - E\left\{Y_3\right\}\right)\left(Y_1 - E\left\{Y_1\right\}\right) & \left(Y_3 - E\left\{Y_3\right\}\right)\left(Y_2 - E\left\{Y_2\right\}\right) & \left(Y_3 - E\left\{Y_3\right\}\right)^2 \end{bmatrix}\right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} = \Sigma$$

Linear Regression Example (n=3)

Error terms are assumed to be independent, with mean 0, constant variance σ^2 :

$$\Rightarrow E\left\{\varepsilon_i\right\}=0 \qquad \sigma^2\left\{\varepsilon_i\right\}=\sigma^2 \qquad \sigma\left\{\varepsilon_i,\varepsilon_j\right\}=0 \quad \forall\, i\neq j$$

$$\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix} \qquad \mathbf{E}\left\{\boldsymbol{\varepsilon}\right\}=\begin{bmatrix}0\\0\\0\end{bmatrix} \qquad \boldsymbol{\sigma}^{\mathbf{2}}\left\{\boldsymbol{\varepsilon}\right\}=\begin{bmatrix}\sigma^2 & 0 & 0\\ 0 & \sigma^2 & 0\\ 0 & 0 & \sigma^2\end{bmatrix}=\sigma^2\mathbf{I}$$

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \qquad \mathbf{E}\left\{\mathbf{Y}\right\}=\mathbf{E}\left\{\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right\}=\mathbf{X}\boldsymbol{\beta}+\mathbf{E}\left\{\boldsymbol{\varepsilon}\right\}=\mathbf{X}\boldsymbol{\beta}$$

$$\boldsymbol{\sigma}^{\mathbf{2}}\left\{\mathbf{Y}\right\}=\boldsymbol{\sigma}^{\mathbf{2}}\left\{\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right\}=\boldsymbol{\sigma}^{\mathbf{2}}\left\{\boldsymbol{\varepsilon}\right\}=\begin{bmatrix}\sigma^2 & 0 & 0\\ 0 & \sigma^2 & 0\\ 0 & 0 & \sigma^2\end{bmatrix}=\sigma^2\mathbf{I}$$

Mean and Variance of Linear Functions of Y

$$\underset{k\times n}{\mathbf{A}} \equiv \text{matrix of fixed constants} \qquad \underset{n\times 1}{\mathbf{Y}} \equiv \text{random vector}$$

$$\underset{k\times 1}{\mathbf{W}} = \mathbf{AY} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \equiv \text{ random vector: } \mathbf{W} = \begin{bmatrix} W_1 \\ \vdots \\ W_k \end{bmatrix} = \begin{bmatrix} a_{11}Y_1 + \ldots + a_{1n}Y_n \\ \vdots \\ a_{k1}Y_1 + \ldots + a_{kn}Y_n \end{bmatrix}$$

$$\mathbf{E}\{\mathbf{W}\} = \begin{bmatrix} E\{W_1\} \\ \vdots \\ E\{W_k\} \end{bmatrix} = \begin{bmatrix} E\{a_{11}Y_1 + \ldots + a_{1n}Y_n\} \\ \vdots \\ E\{a_{k1}Y_1 + \ldots + a_{kn}Y_n\} \end{bmatrix} = \begin{bmatrix} a_{11}E\{Y_1\} + \ldots + a_{1n}E\{Y_n\} \\ \vdots \\ a_{k1}E\{Y_1\} + \ldots + a_{kn}E\{Y_n\} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} E\{Y_1\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} = \mathbf{AE}\{\mathbf{Y}\}$$

$$\boldsymbol{\sigma}^2\{\mathbf{W}\} = \mathbf{E}\left\{\left[\mathbf{AY} - \mathbf{AE}\{\mathbf{Y}\}\right]\left[\mathbf{AY} - \mathbf{AE}\{\mathbf{Y}\}\right]'\right\} = \mathbf{E}\left\{\left[\mathbf{A}\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)\right]\left[\mathbf{A}\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)\right]'\right\} =$$

$$= \mathbf{E}\left\{\left[\mathbf{A}\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)\right]\left[\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)'\mathbf{A}'\right]\right\} = \mathbf{AE}\left\{\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)\left(\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right)'\right\}\mathbf{A}' = \mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{A}'$$

Multivariate Normal Distribution

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \boldsymbol{\mu} = \mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \qquad \boldsymbol{\Sigma} = \boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

Multivariate Normal Density function:

$$f(\mathbf{Y}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \qquad \mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\Rightarrow Y_i \sim N(\mu_i, \sigma_i^2) \quad i = 1, \ldots, n \qquad \sigma\{Y_i, Y_j\} \equiv \sigma_{ij} \quad i \neq j$$

Note, if $\mathbf{A}$ is a (full rank) matrix of fixed constants:

$$\mathbf{W} = \mathbf{AY} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

Simple Linear Regression in Matrix Form

Simple Linear Regression Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \ldots, n$

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

Defining:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \Rightarrow \quad \mathbf{Y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}} \qquad \text{since: } \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} = \mathbf{E}\{\mathbf{Y}\}$$

Assuming constant variance, and independence of error terms ε_i :

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \underset{n \times n}{\mathbf{I}}$$

Further, assuming normal distribution for error terms ε_i : $\quad \mathbf{Y} \sim N\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right)$

Estimating Parameters by Least Squares

Normal equations obtained from: $\frac{\partial Q}{\partial \beta_0}, \frac{\partial Q}{\partial \beta_1}$ and setting each equal to 0:

$$(i) \quad n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \qquad (ii) \quad \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

Note: In matrix form: $\mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \qquad \mathbf{X'Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}$ Defining $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$

$$\Rightarrow \quad \mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{X'Y} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X'X}\right)^{-1} \mathbf{X'Y}$$

Based on matrix form:

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

$$= Y'Y - 2\left(\beta_0 \sum_{i=1}^{n} Y_i + \beta_1 \sum_{i=1}^{n} X_i Y_i\right) + n\beta_0^2 + 2\beta_0\beta_1 \sum_{i=1}^{n} X_i + \beta_1^2 \sum_{i=1}^{n} X_i^2$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}(Q) = \begin{bmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \end{bmatrix} = \begin{bmatrix} -2\sum_{i=1}^{n} Y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} X_i \\ -2\sum_{i=1}^{n} X_i Y_i + 2\beta_0 \sum_{i=1}^{n} X_i + 2\beta_1 \sum_{i=1}^{n} X_i^2 \end{bmatrix} = -2\mathbf{X'Y} + 2\mathbf{X'X}\boldsymbol{\beta} \overset{\text{set}}{=} \mathbf{0} \quad \Rightarrow \quad \mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{X'Y} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X'X}\right)^{-1} \mathbf{X'Y}$$

General Result for fixed symmetric matrix **A** and variable vector **w**: $\frac{\partial}{\partial \mathbf{w}} \mathbf{Aw} = \mathbf{A}' \qquad \frac{\partial}{\partial \mathbf{w}} \mathbf{wAw} = 2\mathbf{Aw}$

Fitted Values and Residuals

$\hat{Y}_i = \hat{\beta}_0 + b_1 X_i \qquad e_i = Y_i - \hat{Y}_i \qquad$ In Matrix form:

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 + \hat{\beta}_1 X_1 \\ \hat{\beta}_0 + \hat{\beta}_1 X_2 \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 X_n \end{bmatrix} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} = \mathbf{PY} \qquad \mathbf{P} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$$

$\mathbf{P}$ is called the "projection" or "hat" matrix, note that $\mathbf{P}$ is idempotent ($\mathbf{PP} = \mathbf{P}$) and symmetric($\mathbf{P} = \mathbf{P'}$):

$$\mathbf{PP} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} = \mathbf{X'I}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} = \mathbf{P} \qquad \mathbf{P'} = \left(\mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\right)' = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} = \mathbf{P}$$

$$\mathbf{e} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{PY} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$$

Note: $\mathbf{E}\left\{\hat{\mathbf{Y}}\right\} = \mathbf{E}\left\{\mathbf{PY}\right\} = \mathbf{PE}\left\{\mathbf{Y}\right\} = \mathbf{PX}\boldsymbol{\beta} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} \qquad \boldsymbol{\sigma}^2\left\{\hat{\mathbf{Y}}\right\} = \mathbf{P}\boldsymbol{\sigma}^2\mathbf{IP'} = \sigma^2\mathbf{P}$

$$\mathbf{E}\left\{\mathbf{e}\right\} = \mathbf{E}\left\{\left(\mathbf{I}-\mathbf{P}\right)\mathbf{Y}\right\} = \left(\mathbf{I}-\mathbf{P}\right)\mathbf{E}\left\{\mathbf{Y}\right\} = \left(\mathbf{I}-\mathbf{P}\right)\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0} \qquad \boldsymbol{\sigma}^2\left\{\mathbf{e}\right\} = \left(\mathbf{I}-\mathbf{P}\right)\boldsymbol{\sigma}^2\mathbf{I}\left(\mathbf{I}-\mathbf{P}\right)' = \sigma^2\left(\mathbf{I}-\mathbf{P}\right)$$

$$\mathbf{s}^2\left\{\hat{\mathbf{Y}}\right\} = MSE\,\mathbf{P} \qquad \mathbf{s}^2\left\{\mathbf{e}\right\} = MSE\left(\mathbf{I}-\mathbf{P}\right)$$

Analysis of Variance

Total (Corrected) Sum of Squares: $SSTO = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$

Note: $\mathbf{Y'Y} = \sum_{i=1}^{n} Y_i^2 \qquad \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = \left(\frac{1}{n}\right)\mathbf{Y'JY} \qquad \underset{n\times n}{\mathbf{J}} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \qquad \Rightarrow SSTO = \mathbf{Y'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY} = \mathbf{Y'}\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$

Defining SSE as the Residual (Error) Sum of Squares:

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \mathbf{e'e} = \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{Y'Y} - \mathbf{Y'X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y} + \hat{\boldsymbol{\beta}}'\mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{Y'Y} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y} = \mathbf{Y'}\left[\mathbf{I} - \mathbf{P}\right]\mathbf{Y}$$

since $\hat{\boldsymbol{\beta}}'\mathbf{X'Y} = \mathbf{Y'X'}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} = \mathbf{Y'PY}$

Defining SSR as the Regression Sum of Squares:

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 SSTO - SSE = \hat{\boldsymbol{\beta}}'\mathbf{X'Y} - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = \mathbf{Y'PY} - \left(\frac{1}{n}\right)\mathbf{Y'JY} = \mathbf{Y'}\left[\mathbf{P} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$$

Note that $SSTO$, SSR, and SSE are all QUADRATIC FORMS: $\mathbf{Y'AY}$ for symmetric and idempotent matrices $\mathbf{A}$

Inferences in Linear Regression

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \quad \Rightarrow \quad \mathbf{E}\{\hat{\boldsymbol{\beta}}\} = (\mathbf{X'X})^{-1}\mathbf{X'E}\{\mathbf{Y}\} = (\mathbf{X'X})^{-1}\mathbf{X'X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

$$\boldsymbol{\sigma}^2\{\hat{\boldsymbol{\beta}}\} = (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{X}(\mathbf{X'X})^{-1} = \boldsymbol{\sigma}^2(\mathbf{X'X})^{-1}\mathbf{X'IX}(\mathbf{X'X})^{-1} = \boldsymbol{\sigma}^2(\mathbf{X'X})^{-1} \qquad \mathbf{s}^2\{\hat{\boldsymbol{\beta}}\} = MSE(\mathbf{X'X})^{-1}$$

where $s^2 = MSE = \dfrac{SSE}{n-2}$

Recall: $$(\mathbf{X'X})^{-1} = \begin{bmatrix} \dfrac{1}{n} + \dfrac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} & -\dfrac{\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ -\dfrac{\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} & \dfrac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \end{bmatrix} \quad \Rightarrow \quad \mathbf{s}^2\{\hat{\boldsymbol{\beta}}\} = \begin{bmatrix} \dfrac{MSE}{n} + \dfrac{\overline{X}^2 MSE}{\sum_{i=1}^{n}(X_i - \overline{X})^2} & -\dfrac{\overline{X}MSE}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ -\dfrac{\overline{X}MSE}{\sum_{i=1}^{n}(X_i - \overline{X})^2} & \dfrac{MSE}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \end{bmatrix}$$

Estimated Mean Response at $X = X_h$:

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \mathbf{X_h}'\hat{\boldsymbol{\beta}} \qquad \mathbf{X_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \qquad s^2\{\hat{Y}_h\} = \mathbf{X_h}'\mathbf{s}^2\{\hat{\boldsymbol{\beta}}\}\mathbf{X_h} = MSE\left(\mathbf{X_h}'(\mathbf{X'X})^{-1}\mathbf{X_h}\right)$$

Predicted New Response at $X = X_h$:

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \mathbf{X_h}'\hat{\boldsymbol{\beta}} \qquad s^2\{\text{pred}\} = MSE\left(1 + \mathbf{X_h}'(\mathbf{X'X})^{-1}\mathbf{X_h}\right)$$