Linear Algebra Review

Yang Feng

http://www.stat.columbia.edu/~yangfeng

Definition of Matrix

- Rectangular array of elements arranged in rows and columns

$$\begin{bmatrix} 16000 & 23 \\ 33000 & 47 \\ 21000 & 35 \end{bmatrix}$$

- A matrix has dimensions
- The dimension of a matrix is its number of rows and columns
- It is expressed as 3×2 (in this case)

Indexing a Matrix

- Rectangular array of elements arranged in rows and columns

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

- A matrix can also be notated

$$\mathbf{A} = [a_{ij}], i = 1, 2; j = 1, 2, 3$$

Square Matrix and Column Vector

- A square matrix has equal number of rows and columns

$$\begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

- A column vector is a matrix with a single column

$$\begin{bmatrix} 4 \\ 7 \\ 10 \end{bmatrix} \qquad \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}$$

- All vectors (row or column) are matrices, all scalars are 1×1 matrices.

Transpose

- The transpose of a matrix is another matrix in which the rows and columns have been interchanged

$$\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 7 & 10 \\ 3 & 4 \end{bmatrix}$$

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

Equality of Matrices

- Two matrices are the same if they have the same dimension and all the elements are equal

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix}$$

$\mathbf{A} = \mathbf{B}$ implies $a_1 = 4, a_2 = 7, a_3 = 3$

Matrix Addition and Substraction

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

Then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{bmatrix}$$

Multiplication of a Matrix by a Scalar

$$\mathbf{A} = \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix}$$

$$k\mathbf{A} = k \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 2k & 7k \\ 9k & 3k \end{bmatrix}$$

Multiplication of two Matrices

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \qquad \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix}$$

$$\begin{array}{cc} & \mathbf{A} \\ \begin{array}{l} \text{Row 1} \\ \text{Row 2} \end{array} & \begin{bmatrix} \boxed{2 \quad 5} \\ 4 \quad 1 \end{bmatrix} \end{array} \qquad \begin{array}{c} \mathbf{B} \\ \begin{bmatrix} 4 & \boxed{\begin{matrix} 6 \\ 8 \end{matrix}} \\ 5 & \end{bmatrix} \\ \text{Col. 1 Col. 2} \end{array} \qquad \begin{array}{cc} & \mathbf{AB} \\ \text{Row 1} & \begin{bmatrix} 33 & 52 \\ & \end{bmatrix} \\ & \text{Col. 1 Col. 2} \end{array}$$

Another Matrix Multiplication Example

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 26 \\ 41 \end{bmatrix}$$

Special Matrices

- If $\mathbf{A} = \mathbf{A}'$, then $\mathbf{A}$ is a symmetric matrix

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

- If the off-diagonal elements of a matrix are all zeros it is then called a diagonal matrix

$$\mathbf{A} = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

Identity Matrix

A diagonal matrix whose diagonal entries are all ones is an identity matrix. Multiplication by an identity matrix leaves the pre or post multiplied matrix unchanged.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

Vector and matrix with all elements equal to one

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix} \quad \mathbf{J} = \begin{bmatrix} 1 & \dots & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & \dots & 1 \end{bmatrix}$$

$$\mathbf{11}' = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & . & . & . & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & \dots & 1 \end{bmatrix} = \mathbf{J}$$

Linear Dependence and Rank of Matrix

Consider

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

and think of this as a matrix of a collection of column vectors.

Note that the third column vector is a multiple of the first column vector.

Linear Dependence

When m scalars $k_1, ..., k_m$ not all zero, can be found such that:

$$k_1\mathbf{A}_1 + ... + k_m\mathbf{A}_m = \mathbf{0}$$

where $\mathbf{0}$ denotes the zero column vector and $\mathbf{A}_i$ is the i^{th} column of matrix $\mathbf{A}$, the m column vectors are called linearly dependent. If the only set of scalars for which the equality holds is $k_1 = 0, ..., k_m = 0$, the set of m column vectors is linearly independent.

In the previous example matrix the columns are linearly dependent.

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

Rank of Matrix

The rank of a matrix is defined to be the maximum number of linearly independent columns in the matrix. Rank properties include

- The rank of a matrix is unique
- The rank of a matrix can equivalently be defined as the maximum number of linearly independent rows
- The rank of an $r \times c$ matrix cannot exceed $min(r, c)$
- The row and column rank of a matrix are equal
- The rank of a matrix is preserved under nonsingular transformations., i.e. Let $\mathbf{A}$ $(n \times n)$ and $\mathbf{C}$ $(k \times k)$ be nonsingular matrices. Then for any $n \times k$ matrix $\mathbf{B}$ we have

$$rank(\mathbf{B}) = rank(\mathbf{AB}) = rank(\mathbf{BC})$$

Inverse of Matrix

- Like a reciprocal

$$6 * 1/6 = 1/6 * 6 = 1$$
$$x\frac{1}{x} = 1$$

- But for matrices

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Example

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

More generally,

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

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

where $D = ad - bc$

Inverses of Diagonal Matrices are Easy

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

then

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

Finding the inverse

- Finding an inverse takes (for general matrices with no special structure)

$$O(n^3)$$

  operations (when n is the number of rows in the matrix)
- We will assume that numerical packages can do this for us in R: solve($\mathbf{A}$) gives the inverse of matrix $\mathbf{A}$

Uses of Inverse Matrix

- Ordinary algebra $5y = 20$
  is solved by $1/5 * (5y) = 1/5 * (20)$
- Linear algebra $\mathbf{AY} = \mathbf{C}$
  is solved by

$$\mathbf{A}^{-1}\mathbf{AY} = \mathbf{A}^{-1}\mathbf{C}, \mathbf{Y} = \mathbf{A}^{-1}\mathbf{C}$$

Example

Solving a system of simultaneous equations

$$\begin{aligned} 2y_1 + 4y_2 &= 20 \\ 3y_1 + y_2 &= 10 \end{aligned}$$

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

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

List of Useful Matrix Properties

$$
\begin{aligned}
\mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A} \\
(\mathbf{A}+\mathbf{B})+\mathbf{C} &= \mathbf{A}+(\mathbf{B}+\mathbf{C}) \\
(\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \\
\mathbf{C}(\mathbf{A}+\mathbf{B}) &= \mathbf{CA}+\mathbf{CB} \\
k(\mathbf{A}+\mathbf{B}) &= k\mathbf{A}+k\mathbf{B} \\
(\mathbf{A}')' &= \mathbf{A} \\
(\mathbf{A}+\mathbf{B})' &= \mathbf{A}'+\mathbf{B}' \\
(\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \\
(\mathbf{ABC})' &= \mathbf{C}'\mathbf{B}'\mathbf{A}' \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{ABC})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A}, (\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'
\end{aligned}
$$

Random Vectors and Matrices

Let's say we have a vector consisting of three random variables

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

The expectation of a random vector is defined as

$$\mathbb{E}(\mathbf{Y}) = \begin{pmatrix} \mathbb{E}(Y_1) \\ \mathbb{E}(Y_2) \\ \mathbb{E}(Y_3) \end{pmatrix}$$

Expectation of a Random Matrix

The expectation of a random matrix is defined similarly

$$\mathbb{E}(\mathbf{Y}) = [\mathbb{E}(Y_{ij})] \qquad i = 1, ...n; j = 1, ..., p$$

Variance-covariance Matrix of a Random Vector

The variances of three random variables $\sigma^2(Y_i)$ and the covariances between any two of the three random variables $\sigma(Y_i, Y_j)$, are assembled in the variance-covariance matrix of $\mathbf{Y}$

$$cov(\mathbf{Y}) = \sigma^2\{\mathbf{Y}\} = \begin{pmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \sigma(Y_1, Y_3) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \sigma(Y_2, Y_3) \\ \sigma(Y_3, Y_1) & \sigma(Y_3, Y_2) & \sigma^2(Y_3) \end{pmatrix}$$

remember $\sigma(Y_2, Y_1) = \sigma(Y_1, Y_2)$ so the covariance matrix is symmetric

Derivation of Covariance Matrix

In vector terms the variance-covariance matrix is defined by

$$\sigma^2\{\mathbf{Y}\} = \mathbb{E}(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))'$$

because

$$\sigma^2\{\mathbf{Y}\} = \mathbb{E}(\begin{pmatrix} Y_1 - \mathbb{E}(Y_1) \\ Y_2 - \mathbb{E}(Y_2) \\ Y_3 - \mathbb{E}(Y_3) \end{pmatrix} \begin{pmatrix} Y_1 - \mathbb{E}(Y_1) & Y_2 - \mathbb{E}(Y_2) & Y_3 - \mathbb{E}(Y_3) \end{pmatrix})$$

Regression Example

- Take a regression example with $n = 3$ with constant error terms $\sigma^2(\epsilon_i)$ and are uncorrelated so that $\sigma^2(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$
- The variance-covariance matrix for the random vector ϵ is

$$\sigma^{\mathbf{2}}(\epsilon) = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

which can be written as $\sigma^{\mathbf{2}}\{\epsilon\} = \sigma^{\mathbf{2}}\ \mathbf{I}$

Basic Results

If $\mathbf{A}$ is a constant matrix and $\mathbf{Y}$ is a random matrix then $\mathbf{W} = \mathbf{AY}$ is a random matrix

$$\begin{gathered}
\mathbb{E}(\mathbf{A}) = \mathbf{A} \\
\mathbb{E}(\mathbf{W}) = \mathbb{E}(\mathbf{AY}) = \mathbf{A}\,\mathbb{E}(\mathbf{Y}) \\
\sigma^2\{\mathbf{W}\} = \sigma^2\{\mathbf{AY}\} = \mathbf{A}\sigma^2\{\mathbf{Y}\}\mathbf{A}'
\end{gathered}$$

Multivariate Normal Density

- Let $\mathbf{Y}$ be a vector of p observations

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_p \end{pmatrix}$$

- Let $\boldsymbol{\mu}$ be a vector of the means of each of the p observations

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_p \end{pmatrix}$$

Multivariate Normal Density

let $\mathbf{\Sigma}$ be the variance-covariance matrix of $\mathbf{Y}$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ . & & & \\ . & & & \\ . & & & \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix}$$

Then the multivariate normal density is given by

$$P(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp[-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})]$$

Matrix Simple Linear Regression

- Nothing new-only matrix formalism for previous results
- Remember the normal error regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, ..., n$$

- Expanded out this looks like

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\ &\dots \\ Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n \end{aligned}$$

- which points towards an obvious matrix formulation.

Regression Matrices

- If we identify the following matrices

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & \\ . & \\ . & \\ 1 & X_n \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

- We can write the linear regression equations in a compact form

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

Regression Matrices

- Of course, in the normal regression model the expected value of each of the ϵ's is zero, we can write $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\beta$
- This is because

$$\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$$

$$\begin{pmatrix} \mathbb{E}(\epsilon_1) \\ \mathbb{E}(\epsilon_2) \\ . \\ . \\ . \\ \mathbb{E}(\epsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

Error Covariance

Because the error terms are independent and have constant variance σ^2

$$\sigma^2\{\boldsymbol{\epsilon}\} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

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

Matrix Normal Regression Model

In matrix terms the normal regression model can be written as

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

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$

Least Square Estimation

If we remember both the starting normal equations that we derived

$$nb_0 + b_1 \sum X_i = \sum Y_i$$
$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i$$

and the fact that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Least Square Estimation

Then we can see that these equations are equivalent to the following matrix operations

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

with

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

with the solution to this equation given by

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

when $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

Fitted Value

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

Because:

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ . \\ . \\ . \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ . \\ . \\ . \\ b_0 + b_1 X_n \end{pmatrix}$$

Fitted Values, Hat Matrix

plug in

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

We have

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

or

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

where

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

is called the hat matrix.
Property of hat matrix $\mathbf{H}$:

1. symmetric
2. idempotent: $\mathbf{H}\mathbf{H} = \mathbf{H}$.

Residuals

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

Then

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

The matrix $\mathbf{I} - \mathbf{H}$ is also symmetric and idempotent.
The variance-covariance matrix of $\mathbf{e}$ is

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

And we can estimate it by

$$\mathbf{s}^2\{\mathbf{e}\} = MSE(\mathbf{I} - \mathbf{H})$$

Analysis of Variance Results

$$SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

We know

$$\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$$

and $\mathbf{J}$ is the matrix with entries all equal to 1. Then we have

$$\frac{(\sum Y_i)^2}{n} = \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

As a result:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

Analysis of Variance Results

Also,

$$SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

can be represented as

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

Notice that $\mathbf{H1} = \mathbf{1}$, then $(\mathbf{I}-\mathbf{H})\mathbf{J} = \mathbf{0}$
Finally by similarly reasoning,

$$SSR = ([\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y})'([\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y}) = \mathbf{Y}'[\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y}$$

Easy to check that

$$SSTO = SSE + SSR$$

Sums of Squares as Quadratic Forms

When $n = 2$, an example of quadratic forms:

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2$$

can be expressed as matrix term as

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$$

In general, a quadratic term is defined as :

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} Y_i Y_j$$

where $A_{ij} = A_{ji}$
Here, **A** is a symmetric $n \times n$ matrix , the matrix of the quadratic form.

Quadratic forms for ANOVA

$$SSTO = \mathbf{Y}'[\mathbf{I} - \frac{1}{n}\mathbf{J}]\mathbf{Y}$$

$$SSE = \mathbf{Y}'[\mathbf{I} - \mathbf{H}]\mathbf{Y}$$

$$SSR = \mathbf{Y}'[\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y}$$

Inference in Regression Analysis

- Regression Coefficients: The variance-covariance matrix of **b** is

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

- Mean Response: To estimate the mean response at X_h, define $\mathbf{X}_h = \begin{pmatrix} 1 \\ X_h \end{pmatrix}$ Then

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

And the variance-covariance matrix of $\hat{Y}_h$ is

$$\sigma^2\{\hat{Y}_h\} = \mathbf{X}_h'\sigma^2\{\mathbf{b}\}\mathbf{X}_h = \sigma^2\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h$$

- Prediction of New Observation:

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