Matrix Approach to Simple Linear Regression Analysis

5.1-5.4 Matrices

Definition. *Matrix*

1. ***Matrix*** $A = [a_{ij}]$

2. ***Vector****, matrix containing only 1 column is called a* ***column vector****, or simply a* ***vector***

3. ***Matrix transpose*** $A' = [a_{ij}]' = [a_{ji}]$

4. ***Matrix equivalence*** *2 matricies are equivalent if they have same dimension and if all corresponding elements are equal*

5. ***Matricies in Regression*** *vector* $\mathbf{Y}_{n\times 1}$ *and vector* $\mathbf{X}_{n\times 2}$ *(also referred to as the design matrix)*

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

6. ***Matrix addition and subtraction*** *Requires having same dimension*

$$\mathbf{A}_{r\times c} + \mathbf{B}_{r\times c} = [a_{ij} + b_{ij}] \qquad \mathbf{A}_{r\times c} - \mathbf{B}_{r\times c} = [a_{ij} - b_{ij}]$$

Also

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

7. ***Regression model as Matrices***

$$Y_i = \mathbb{E}(Y_i) + \epsilon_i \quad i = 1, \cdots, n$$

can be written as

$$\mathbf{Y}_{n\times 1} = \mathbb{E}(\mathbf{Y})_{n\times 1} + \epsilon_{n\times 1}$$

be interpreted as observation vector $\mathbf{Y}$ *equals the sum of 2 vectors, a vector containing the expected value and another containing the error terms*

8. ***Scalar Multiplication***

$$k\mathbf{A} = \mathbf{A}k = [ka_{ij}] \qquad \text{for some scalar } k$$

9. ***Matrix Multiplication*** *Only defined if number of columns in* $\mathbf{A}$ *equals number of rows in* $\mathbf{B}$

$$\mathbf{A_{r\times c}B_{c\times s}} = \left[\sum_{k=1}^{c} a_{ik}b_{kj}\right]_{r\times s} \qquad i = 1, \cdots, r \quad j = 1, \cdots, s$$

with

$$\mathbf{AB} \neq \mathbf{BA}$$

in regression analysis we often find

$$\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

$$\mathbf{Y'Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \left[\sum Y_i^2\right] = \sum Y_i^2$$

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

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

10. ***Types of matricies***

    (a) ***Symmetric matrices*** *If* $\mathbf{A} = \mathbf{A}'$*, then* $\mathbf{A}$ *is symmetric. A symmetric matrix is necessarily square. (Note* $\mathbf{XX}'$ *is symmetric)*

    (b) ***Diagonal matrix*** *is a square matrix whose off-diagonal elements are all zeros.*

    (c) ***Identity matrix*** *or unit matrix is denoted by* $\mathbf{I}$*, a diagonal matrix whose elements on the main diagonals are all 1s with property*

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

    (d) ***Scalar matrix*** *is a diagonal matrix whose main-diagonal elements are the same. So it can be expressed as* $k\mathbf{I}$ *where* k *is a scalar*

(e) ***Vector and Matrix with all elements unity*** *A column vector with elements 1 is denoted by* $\mathbf{1}$*. A square matrix with all elements 1 is denoted by* $\mathbf{J}$

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

with properties

$$\mathbf{1}'\mathbf{1}_{1\times 1} = n \qquad \mathbf{1}\mathbf{1}'_{n\times n} = \mathbf{J}_{n\times n}$$

(f) ***Zero vector*** *is a vector containing only zeros*

$$\mathbf{0}_{r\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

11. ***Linear Dependence***
Define a set of c columns $\mathbf{C_1}, \cdots, \mathbf{C_n}$ *in an* $r \times c$ *matrix is linearly dependent if one vector can be expressed as linear combination of others. If no vector in the set can be so expressed, we define the set of vectors to be linearly independent. Equivalently,*

When c scaslars $k_1, \cdots, k_c$*, not all zero, can be found such that*

$$k_1\mathbf{C_1} + k_2\mathbf{C_2} + \cdots + k_c\mathbf{C_c} = \mathbf{0}$$

where $\mathbf{0}$ *denotes zero column vector, the c column vectors are* ***linearly dependent****. If the only set of scalars for which the equality holds is* $k_1 = 0, \cdots, k_c = 0$*, the set of c column vectors is* ***linearly independent***

12. ***Rank of Matrix*** *is defined to be maximum number of linearly independent columns in the matrix. Rank of a matrix is unique and can be equivalently defined as the maximum number of linearly independent rows. It follows that*

    (a) *rank of* $r \times c$ *matrix cannot exceed* $\min\{r, c\}$.

    (b) *when a matrix is product of 2 matrices, its rank cannot exceed the smaller of the two ranks for the matrices beging multiplied.*

$$\mathbf{C} = \mathbf{AB} \rightarrow rank\mathbf{C} \leq \min\{rank\mathbf{A}, rank\mathbf{B}\}$$

13. ***Inverse of a Matrix*** *THe inverse of a matrix is denoted by* $\mathbf{A^{-1}}$ *such that*

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

where $\mathbf{I}$ *is the identity matrix.*

(a) *Inverse of matrix is only defined for square matrices.*

(b) *If a square matrix does have an inverse, it is unique. Even so many square matrices do not have inverses.*

(c) *An inverse of $r \times r$ matrix exists if the rank of the matrix is r. Such a matrix is **nonsingular** or **full rank**. An $r \times r$ matrix with rank less than r does not have an inverse. Note **singular** matricies has determinant of zero, and thus no inverses.*

(d) *The inverse of an $r \times r$ matrix of full rank also has rank r*

To find the inverse of matrices

$$A_{2\times 2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A_{2\times 2}^{-1} = \begin{bmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{bmatrix} \quad \textit{where} \quad D = ad - bc$$

here D is the determinant
In regression analysis, we have

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

with determinant

$$D = n \sum X_i^2 - (\sum X_i)^2 = n \left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] = n \sum (X_i - \overline{X})^2$$

so we can find the inverse

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{n \sum (X_i - \overline{X})^2} & \frac{-\sum X_i}{n \sum (X_i - \overline{X})^2} \\ \frac{-\sum X_i}{n \sum (X_i - \overline{X})^2} & \frac{n}{n \sum (X_i - \overline{X})^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} + \frac{\overline{X}^2}{S_{XX}} & \frac{-\overline{X}}{S_{XX}} \\ \frac{-\overline{X}}{S_{XX}} & \frac{1}{S_{XX}} \end{bmatrix}$$

Note

$$\frac{\sum X_i^2}{n \sum (X_i - \overline{X})^2} = \frac{\sum X_i^2 - n\overline{X}^2 + n\overline{X}^2}{n \sum (X_i - \overline{X})^2} = \frac{1}{n} + \frac{\overline{X}^2}{S_{XX}}$$

14. ***Properties of Matrices***

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

15. ***Random vectors and matrices*** *contains element that are random variables. Expected value of a random vector is a vector whose elements are the expected values of the random variables that are the elements of the random vectors*

$$\underset{n\times p}{\mathbb{E}(\mathbf{Y})}=[\mathbb{E}(Y_{ij})] \qquad i=1,\cdots,n \quad j=1,\cdots,p$$

16. ***Variance-Covariance Matrix of random vector***
variances $\sigma^2\{Y_i\}$ and covariances between any two of the random variables $\sigma\{Y_i,Y_j\}$ are assembled in the variance-covariance matrix of vector $\underset{n\times 1}{\mathbf{Y}}$, denoted by $\sigma^{\mathbf{2}}\{\mathbf{Y}\}$

$$\underset{n\times n}{\sigma^{\mathbf{2}}\{\mathbf{Y}\}}=\mathbb{E}\left\{[\mathbf{Y}-\mathbb{E}(\mathbf{Y})][\mathbf{Y}-\mathbb{E}(\mathbf{Y})]'\right\}=\begin{bmatrix}\sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \cdots & \sigma\{Y_1,Y_n\}\\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \cdots & \sigma\{Y_2,Y_n\}\\ \vdots & \vdots & & \vdots\\ \sigma\{Y_n,Y_1\} & \sigma\{Y_n,Y_2\} & \cdots & \sigma^2\{Y_n\}\end{bmatrix}$$

Note $\sigma^{\mathbf{2}}\{\mathbf{Y}\}$ is a symmetric matrix by $\sigma\{Y_i,Y_j\}=\sigma\{Y_j,Y_i\}$ for all $i\neq j$
In regression analysis, if error terms have constant variance $\sigma^2\{\epsilon_i\}=\sigma^2$ and are uncorrelated $\sigma\{\epsilon_i,\epsilon_j\}=0$ for $i\neq j$, then we have

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

17. ***Expectation and variance-covariance matrix of constant Matrix transformation***
For a random vector $\mathbf{W}$ *obtained by premultiplying the random vector* $\mathbf{Y}$ *by a constant matrix* $\mathbf{A}$

$$\mathbf{W} = \mathbf{AY}$$

we have

$$\begin{aligned}\mathbb{E}\{\mathbf{A}\} &= \mathbf{A} \\ \mathbb{E}\{\mathbf{W}\} &= \mathbb{E}\{\mathbf{AY}\} = \mathbf{A}\mathbb{E}\{\mathbf{Y}\} \\ \sigma^2\{\mathbf{W}\} &= \boldsymbol{\sigma^2}\{\mathbf{AY}\} = \mathbf{A}\boldsymbol{\sigma^2}\{\mathbf{Y}\}\mathbf{A}'\end{aligned}$$

5.9 SLR model in Matrix Terms

Definition. ***Simple Linear Regression in Matrix Terms***

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, \cdots, n$$

can be written as

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}} \text{ where } \mathbb{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$$

where

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{n\times 1}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Regression model assumes $\mathbb{E}\{\epsilon_i\} = 0$, *and* $\sigma^2\{\epsilon_i\} = \sigma^2$ *and that* ϵ_i *are indepedent normal random variables. Equivalently,*

$$\mathbb{E}\{\boldsymbol{\epsilon}\} = \underset{n\times 1}{\mathbf{0}} \qquad \boldsymbol{\sigma^2}\{\boldsymbol{\epsilon}\} = \sigma^2\underset{n\times n}{\mathbf{I}}$$

In summary, normal error regression model in matrix term is

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

$\boldsymbol{\epsilon}$ *is a vector of independent normal random variables with* $\mathbb{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}$ *and* $\boldsymbol{\sigma^2}\{\boldsymbol{\epsilon}\} = \sigma^2\mathbf{I}$

The Gauss Markov theorem can be summarized as follows

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

5.10-11 Least Square Estimation, residual, and fitted value

1. **Normal Equation**

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i = \sum Y_i$$
$$\hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 = \sum X_i Y_i$$

can be written as

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}}\ \underset{2\times 1}{\hat{\boldsymbol{\beta}}} = \underset{2\times 1}{\mathbf{X}'\mathbf{Y}}$$

Proof.

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \mathbf{X}'\mathbf{Y}$$

□

2. **Estimated Regression Coefficients**

To derive normal equation, we minimize

$$Q = \sum [Y_i - (\beta_0 + \beta_1 X_i)]^2$$

which has matrix form of

$$\begin{aligned} Q &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Last step since $\mathbf{Y}'\mathbf{X}\boldsymbol{\beta}$ is a 1×1 matrix so equal to its transpose. To minimize $\hat{\boldsymbol{\beta}}$ we take derivative

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

It follows that

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

yields normal equation.

To obtain estimated regression coefficients from normal equation, we premultiply both sides by the inverse of $\mathbf{X}'\mathbf{X}$

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

3. **Fitted value**

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \qquad \text{where} \qquad \hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

We can express matrix result for $\hat{\mathbf{Y}}$ by using expression for estimated regression coefficient

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

Or equivalently,

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y} \qquad \text{where} \qquad \underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$$

Fitted value $\hat{Y}_i$ can be expressed as a linear combinations of response variable observations Y_i, with the coefficients being elements of the **Hat Matrix $\hat{\mathbf{Y}}$**. Hat Matrix has properties

(a) $\mathbf{H}$ is symmetric

(b) $\mathbf{H}$ is idempotent, i.e.

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

4. **Residuals**

$$\boldsymbol{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

Note e_i, like the fitted value $\hat{Y}_i$ can be expressed as linear combinations of response variable observations Y_i, using the hat matrix

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

Note $\mathbf{I} - \mathbf{H}$ is also symmetric and idempotent. The variance-covariance matrix of vector of residuals can be represented as

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

which can be estimated by

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

Proof.

$$\begin{aligned}\sigma^2\{\mathbf{e}\} &= \sigma^2\{(\mathbf{I}-\mathbf{H})\mathbf{Y}\}\\ &= (\mathbf{I}-\mathbf{H})\sigma^2\{Y\}(\mathbf{I}-\mathbf{H})'\\ &= (\mathbf{I}-\mathbf{H})\sigma^2\{\boldsymbol{\epsilon}\}(\mathbf{I}-\mathbf{H})'\\ &= (\mathbf{I}-\mathbf{H})\sigma^2\mathbf{I}(\mathbf{I}-\mathbf{H})'\\ &= \sigma^2(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H})\\ &= \sigma^2(\mathbf{I}-\mathbf{H})\end{aligned}$$

By symmetric and idempotent property of $\mathbf{I}-\mathbf{H}$ □

5.12 Analysis of Variance Results

1. **Sum of Squares**