Chapter 5: Matrix approach to simple linear regression analysis

Need to understand matrix algebra for multiple regression.

5.1 Matrices

What is a matrix?

A matrix is a rectangular array of elements arranged in rows and columns.

Example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Dimension – Size of matrix: # rows × # columns = r×c

Example: 2×3

Symbolic representation of a matrix:

Example:

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

where a_{ij} is the row i and column j element of **A**

a_{11}=1 from the above example

Notice that the matrix **A** is in bold.

a_{11} is often called the "(1,1) element" of **A**, a_{12} is called the "(1,2) element" of **A**,…

Example: r×c matrix

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

Example: Square matrix is r×c where r=c

Vector – a r×1 (column vector) or 1×c (row vector) matrix – special case of a matrix

Example: Symbolic representation of a 3×1 column vector

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Example:

$$\mathbf{Y} = \begin{bmatrix} 3.10 \\ 2.30 \\ 3.00 \\ \vdots \\ 2.20 \\ 1.60 \end{bmatrix}$$

Transpose: Interchange the rows and columns of a matrix or vector

Example:

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

A is 2×3 and **A′** is 3×2

Example:

$$\mathbf{Y}' = \begin{bmatrix} 3.10 & 2.30 & 3.00 & \cdots & 2.20 & 1.60 \end{bmatrix}$$

Equality of matrices – Two matrices are equal if all of their elements are equal.

5.2 Matrix addition and subtraction

Add or subtract the corresponding elements of matrices with the same dimension.

Example:

Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 10 & -1 \\ 5 & 5 & 8 \end{bmatrix}$. Then

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} 0 & 12 & 2 \\ 9 & 10 & 14 \end{bmatrix} \text{ and } \mathbf{A}-\mathbf{B} = \begin{bmatrix} 2 & -8 & 4 \\ -1 & 0 & -2 \end{bmatrix}.$$

Example: Simple linear regression model

$Y_i = E(Y_i) + \varepsilon_i$ for i=1,…,n can be represented as

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

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \; E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}, \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

5.3 Matrix multiplication

Scalar - 1×1 matrix

Example: Matrix multiplied by a scalar

$$c\mathbf{A} = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{bmatrix}$$ where c is a scalar

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and c=2. Then $2\mathbf{A} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$.

Multiplying two matrices

Suppose you want to multiply the matrices **A** and **B**; i.e., **A**∗**B** or **AB**. In order to do this, you need the number of columns of **A** to be the same as the number of rows as **B**. For example, suppose **A** is 2×3 and **B** is 3×10. You can multiply these matrices. However if **B** is 4×10 instead, these matrices could NOT be multiplied.

The resulting dimension of **C**=**AB**

1. The number of rows of **A** is the number of rows of **C**.
2. The number of columns of **B** is the number of rows of **C**.
3. In other words, $\underset{w\times y}{\mathbf{C}} = \underset{w\times z}{\mathbf{A}}\,\underset{z\times y}{\mathbf{B}}$ where the dimension of the matrices are shown below them.

How to multiply two matrices – an example

Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$. Notice that **A** is 2×3 and **B** is 3×2 so **C**=**AB** can be done.

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1*3+2*1+3*0 & 1*0+2*2+3*1 \\ 4*3+5*1+6*0 & 4*0+5*2+6*1 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 7 \\ 17 & 16 \end{bmatrix}$$

The “cross product” of the rows of **A** and the columns of **B** are taken to form **C**

In the above example, **D**=**BA**≠**AB** where **BA** is:

$$\mathbf{BA} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 3*1+0*4 & 3*2+0*5 & 3*3+0*6 \\ 1*1+2*4 & 1*2+2*5 & 1*3+2*6 \\ 0*1+1*4 & 0*2+1*5 & 0*3+1*6 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 6 & 9 \\ 9 & 12 & 15 \\ 4 & 5 & 6 \end{bmatrix}$$

In general for a 2×3 matrix times a 3×2 matrix:

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

Example:

$$\mathbf{X} = \begin{bmatrix} 1 & 3.04 \\ 1 & 2.35 \\ 1 & 2.70 \\ \vdots & \vdots \\ 1 & 2.28 \\ 1 & 1.88 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} 3.10 \\ 2.30 \\ 3.00 \\ \vdots \\ 2.20 \\ 1.60 \end{bmatrix}$$

Find $\mathbf{X'X}$, $\mathbf{X'Y}$, and $\mathbf{Y'Y}$

Notes:

1.
$$\mathbf{Y'Y} = [y_1, y_2, \ldots, y_n]\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} y_i^2$$

2.
$$\mathbf{X'Y} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_{i1} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \end{bmatrix}$$ since $x_{11}=\ldots=x_{n1}=1$

3. $\mathbf{X'X}$
$$= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \end{bmatrix}\begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{n2} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i2} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}^2 \end{bmatrix}$$

5.4 Special types of matrices

Symmetric matrix: If $\mathbf{A}=\mathbf{A}'$, then $\mathbf{A}$ is symmetric.

Example: $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$, $\mathbf{A}' = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$

Diagonal matrix: A square matrix whose “off-diagonal” elements are 0.

Example: $\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$

Identity matrix: **A** diagonal matrix with 1’s on the diagonal.

Example:

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

Note that "**I**" (the letter **I**, not the number one) usually denotes the identity matrix.

Vector and matrix of 1's

A column vector of 1's:

$$\underset{r\times 1}{\mathbf{1}} = \underset{r\times 1}{\mathbf{j}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

A matrix of 1's:

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

Notes:

1. $\underset{r\times 1}{\mathbf{j}'}\underset{r\times 1}{\mathbf{j}} = r$
2. $\underset{r\times 1}{\mathbf{j}}\underset{r\times 1}{\mathbf{j}'} = \underset{r\times r}{\mathbf{J}}$
3. $\underset{r\times r}{\mathbf{J}'}\underset{r\times r}{\mathbf{J}} = r\underset{r\times r}{\mathbf{J}}$

Vector of 0's:

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

5.5 Linear dependence and rank of matrix

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 6 \\ 3 & 4 & 12 \\ 5 & 6 & 18 \end{bmatrix}$. Think of each column of **A** as a vector; i.e., $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]$. Note that $3\mathbf{A}_2=\mathbf{A}_3$. This means the columns of **A** are "linearly dependent."

Formally, a set of column vectors are **linearly dependent** if there exists constants $\lambda_1, \lambda_2, \ldots, \lambda_n$ (not all zero) such that $\lambda_1\mathbf{A}_1+\lambda_2\mathbf{A}_2+\ldots+\lambda_c\mathbf{A}_c=\mathbf{0}$. A set of column vectors are **linearly independent** if $\lambda_1\mathbf{A}_1+\lambda_2\mathbf{A}_2+\ldots+\lambda_c\mathbf{A}_c=\mathbf{0}$ only for $\lambda_1=\lambda_2=\ldots=\lambda_c=\mathbf{0}$.

The rank of a matrix is the maximum number of linearly independent columns in the matrix.

rank(**A**)=2

5.6 Inverse of a matrix

Note that the inverse of a scalar, say b, is b^{-1}. For example, the inverse of b=3 is $3^{-1}=1/3$. Also, $b*b^{-1}=1$. In matrix algebra, the inverse of a matrix is another matrix. For example, the inverse of **A** is $\mathbf{A}^{-1}$, and $\mathbf{A}*\mathbf{A}^{-1}=\mathbf{A}^{-1}*\mathbf{A}=\mathbf{I}$. Note that A must be a square matrix.

Example: $\mathbf{A}=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$, $\mathbf{A}^{-1}=\begin{bmatrix}-2 & 1\\ 1.5 & -0.5\end{bmatrix}$

Check: $\mathbf{AA}^{-1}=\begin{bmatrix}1*(-2)+2*1.5 & 1*1+2*(-0.5)\\ 3*(-2)+4*1.5 & 3*1+4*(-0.5)\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

Finding the inverse

For a general way, see a matrix algebra book (such as: Kolman, 1988). For a 2×2 matrix, there is a simple formula. Let $\mathbf{A}=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}$.

Then $\mathbf{A}^{-1}=\dfrac{1}{a_{11}a_{22}-a_{12}a_{21}}\begin{bmatrix}a_{22} & -a_{12}\\ -a_{21} & a_{11}\end{bmatrix}$.

Verify $\mathbf{AA}^{-1}=\mathbf{I}$ on your own.

Example: College and HS GPA (GPA_example5.sas)

Remember that $\mathbf{X}=\begin{bmatrix}1 & 3.04\\ 1 & 2.35\\ 1 & 2.70\\ \vdots & \vdots\\ 1 & 2.28\\ 1 & 1.88\end{bmatrix}$ and $\mathbf{Y}=\begin{bmatrix}3.10\\ 2.30\\ 3.00\\ \vdots\\ 2.20\\ 1.60\end{bmatrix}$

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

Note that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}=\underset{2\times 1}{\mathbf{b}}=\begin{bmatrix}b0\\ b1\end{bmatrix}$!!!

Read "Uses of Inverse Matrix" on p. 192-193 of KNN.

Look at some R codes: Chapter5_Ch6_PR05.R

5.7 Some basic theorems of matrices (Page 193)

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
3. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
4. $\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$
5. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$
6. $(\mathbf{A}')' = \mathbf{A}$
7. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
8. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
9. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$
10. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
11. $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$
12. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
13. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$

5.8 Random vectors and matrices

A random vector or random matrix contains elements that are random variables.

Example: Simple linear regression model

$Y_i = E(Y_i) + \varepsilon_i$ for $i=1,\ldots,n$ can be represented as

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

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

Expectation of random vector or matrix: Find the expected value of the individual elements

Example:

$$E(\mathbf{Y}) = E\left(\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}\right) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

$$E(\boldsymbol{\varepsilon}) = E\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}\right) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$ since we assume $\varepsilon_i \sim N(0,\sigma^2)$

Variance-covariance matrix of a random vector

Let Z_1 and Z_2 be random variables. Remember that $\sigma^2\{Z_1\}=E(Z_1-\mu_1)^2$ where $E(Z_1)=\mu_1$.

The covariance of Z_1 and Z_2 is defined as $s\{Z_1,Z_2\} = E[(Z_1-\mu_1)(Z_2-\mu_2)]$. The covariance measures the relationship between Z_1 and Z_2.

Notes:

- $\sigma\{Z_1,Z_1\}=E[(Z_1-\mu_1)(Z_1-\mu_1)]=E(Z_1-\mu_1)^2= \sigma^2\{Z_1\}$
- If Z_1 and Z_2 are independent, then $\sigma\{Z_1,Z_2\}=0$.

A variance-covariance matrix (most often just called the covariance matrix) is a matrix whose elements are the variances and covariances of random variables.

Example: Simple linear regression model

Y is a n×1 random vector

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

$$\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}$$

Note that all covariance matrices are symmetric!

Some basic theorems

Let **W** = **AY** where **Y** is a random vector and **A** is a matrix of constants (no random variables in it).

The following results follow:

1. E(**A**) = **A**

   Remember this is like saying E(3) = 3

2. E(**W**) = E(**AY**) = **A**E(**Y**)

   Again, this is like saying $E(3Y_1)=3E(Y_1)$

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

   You may have seen before that $\sigma^2(aY_1) = a^2\,\sigma^2\{Y_1\}$ where a is a constant.

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$. Let **W**=**AY**, then $\mathbf{W} = \begin{bmatrix} Y_1 - Y_2 \\ Y_1 \end{bmatrix}$.

What is E(W) and $\boldsymbol{\sigma}^2\{\mathbf{W}\}$?

5.9 Simple linear regression model in matrix terms

$Y_i=\beta_0+\beta_1X_i+\varepsilon_i$

where $\varepsilon_i\sim$ independent $N(0,\sigma^2)$ and i=1,…,n

The model can be rewritten as

$Y_1=\beta_0+\beta_1X_1+\varepsilon_1$
$Y_2=\beta_0+\beta_1X_2+\varepsilon_2$
$Y_3=\beta_0+\beta_1X_3+\varepsilon_3$
$\vdots$
$Y_n=\beta_0+\beta_1X_n+\varepsilon_n$

In matrix terms, let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ \mathbf{X}=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix},\ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, which is

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

Note that $E(\mathbf{Y}) = E(\mathbf{X\beta} + \boldsymbol{\varepsilon}) = E(\mathbf{X\beta}) = \mathbf{X\beta}$ since $E(\boldsymbol{\varepsilon}) = \mathbf{0}$

5.10 Least squares estimation of regression parameters

From Chapter 1: The least squares method tries to find the b_0 and b_1 such that SSE $= \Sigma(Y-\hat{Y})^2 = \Sigma(\text{residual})^2$ is minimized. Formulas for b0 and b1 are derived using calculus.

It can be shown that b_0 and b_1 can be found from solving the "normal equations":

$$-2\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i) = 0 \;\Rightarrow\; \sum_{i=1}^{n}Y_i = nb_0 + b_1\sum_{i=1}^{n}X_i \qquad (1a)$$

$$-2\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)X_i = 0 \;\Rightarrow\; \sum_{i=1}^{n}X_iY_i = b_0\sum_{i=1}^{n}X_i + b_1\sum_{i=1}^{n}X_i^2 \qquad (2a)$$

The normal equations can be rewritten as $\mathbf{X'Xb}=\mathbf{X'Y}$, where $\underset{2\times 1}{\mathbf{b}} = \begin{bmatrix} b0 \\ b1 \end{bmatrix}$. Recall in Section 5.3, it was shown that

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_iY_i \end{bmatrix}$$ and

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

Therefore, we have a way to find **b** using matrix algebra:

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

Example: Toluca example (Page 200).

5.11 Fitted values and residuals

Let $\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$, Then $\hat{\mathbf{Y}} = \mathbf{Xb}$.

Hat matrix: $\hat{\mathbf{Y}} = \mathbf{Xb} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}X'Y = HY$, where **H**=**X**(**X′X**)$^{-1}$**X′** is the hat matrix.

The matrix is symmetric and has the special property (called idempotency): **H H= H.**

We will use this in Chapter 9 to measure the influence of observations on the estimated regression line.

Residuals:

Let $\mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix}$, then

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

Covariance matrix of the residuals:

$\sigma^2\{\mathbf{e}\} = \sigma^2(\mathbf{I}-\mathbf{H})$ (see p. 204 for derivation) and is estimated by $\mathbf{s}^2\{\mathbf{e}\} = MSE(\mathbf{I} - \mathbf{H})$.

5.12 Analysis of variance results

Sums of Squares

From Chapter 2: $SSTO = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, $SSE = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$, and

$$SSR = SSTO - SSE = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

These can be rewritten using matrices:

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

$$= \begin{bmatrix} Y_1 & \cdots & Y_n \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} - \frac{1}{n}\begin{bmatrix} Y_1 & \cdots & Y_n \end{bmatrix}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} Y_i & \cdots & \sum_{i=1}^{n} Y_i \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$SSE = \mathbf{e'e} = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y'Y} - \mathbf{B'X'Y}$$

$$\mathrm{SSR} = \mathbf{b'X'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY}$$

5.13 Inferences in Regression Analysis

Covariance matrix of **b**:

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

Estimated covariance matrix of **b**:

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

Estimated variance used in the C.I. for $E(Y_h)$ (mean response):

Define the vector

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix},$$

The fitted value in matrix notation is $\hat{Y}_h = \mathbf{X}_h'\mathbf{b}$, then

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

Estimated variance used in the P.I. for $Y_{h(new)}$ (predcition of new observation):

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

R tips: Remember that the C.l. and P.I. can also be found using the built in functions in R. Also, the estimated covariance matrix of **b** can be found using the VCOV function.

```
#######################################
## You can use some built-in R functions
Data <-
read.table(file="C:/Hongmei/Teaching/stat350/DataSet/CH06PR0
5.txt", header=FALSE)

# Give names to the columns of data set
colnames(Data) <- c("Y", "X1", "X2")
fit <- lm(Y ~ X1, data=Data)
# or use fit <- lm(Data$Y ~ Data$X1)
fit
anova(fit)

# The estimated variance-covariance matrix for b
cov.b <- vcov(fit)
cov.b
```