Chapter 5

Matrix Approach to Simple Linear Regression

Instructor: Li C.X.

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

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

Special Matrix Types

Identity and Scalar Matrix:

$$\underset{3\times 3}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} k & 0 & 0 & 0 \\ 0 & k & 0 & 0 \\ 0 & 0 & k & 0 \\ 0 & 0 & 0 & k \end{bmatrix} = k \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = k \underset{4\times 4}{\mathbf{I}}$$

1-Vector and matrix and zero-vector:

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

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

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):

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

Idempotent matrix and trace

- The **trace** of an n-by-n square matrix $\mathbf{A}$

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}'), \qquad \mathrm{tr}\Big(\underset{m\times n}{\mathbf{A}}\ \underset{n\times m}{\mathbf{B}}\Big) = \mathrm{tr}\Big(\underset{n\times m}{\mathbf{B}}\ \underset{m\times n}{\mathbf{A}}\Big)$$

- The matrix $\mathbf{A}$ is idempotent if and only if $\mathbf{A}^2 = \mathbf{A}$.
  - $\mathbf{A}$ is idempotent $\Rightarrow \mathbf{A}^n = \mathbf{A},\ n=1,2,3,\ldots$
- An idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1.

$$\lambda \mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{A}^2\mathbf{x} = \lambda \mathbf{A}\mathbf{x} = \lambda^2 \mathbf{x} \quad \Rightarrow \quad (\lambda - \lambda^2)\,\mathbf{x} = \mathbf{0}$$

$$\Rightarrow \quad (\lambda - \lambda^2) = \mathbf{0} \quad \Rightarrow \quad \lambda = 0 \text{ or } \lambda = 1$$

Idempotent matrix

- For an idempotent matrix $\mathbf{A}$, rank($\mathbf{A}$) = trace($\mathbf{A}$)， the number of non-zero eigenvalues of A.

- Suppose $\mathbf{A}_1$ and $\mathbf{A}_2$ are idempotent n-by-n matrices, then

  (1) $\mathbf{A}_1 + \mathbf{A}_2$ is idempotent $\Leftrightarrow$ $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1 = \mathbf{0}$

  (2) $\mathbf{A}_1 - \mathbf{A}_2$ is idempotent $\Leftrightarrow$ $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1 = \mathbf{A}_2$

- **Corollary:** $\mathbf{A}$ is idempotent $\Rightarrow$ $\mathbf{I} - \mathbf{A}$ is idempotent.

Random Vectors and Matrices

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

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

- The expectation of a random vector is defined as

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

$$\text{In general: } \underset{n\times p}{E\{\mathbf{Y}\}} = \left[E\{Y_{ij}\} \right] \quad i = 1,\ldots,n;\ j = 1,\ldots,p$$

Variance-covariance Matrix of a Random Vector

$$\operatorname{var}\{\mathbf{Y}\} = E\left\{\left[\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right]\left[\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}\right]'\right\}$$

$$= \mathbf{E}\left\{\begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ Y_3 - E\{Y_3\} \end{bmatrix}\begin{bmatrix} Y_1 - E\{Y_1\} & Y_2 - E\{Y_2\} & Y_3 - E\{Y_3\} \end{bmatrix}\right\}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

The variance-covariance matrix is symmetric.

Covariance Matrix of Two Random Vectors

$$\operatorname{cov}\left\{\underset{m\times 1}{\mathbf{X}},\underset{n\times 1}{\mathbf{Y}}\right\}=E\left\{\left[\mathbf{X}-\mathbf{E}\left\{\mathbf{X}\right\}\right]\left[\mathbf{Y}-\mathbf{E}\left\{\mathbf{Y}\right\}\right]'\right\}$$

$$=E\left\{\begin{bmatrix} X_1-E\left\{X_1\right\} \\ \cdots \\ X_m-E\left\{X_m\right\} \end{bmatrix}\begin{bmatrix} Y_1-E\left\{Y_1\right\} & \cdots & Y_n-E\left\{Y_n\right\} \end{bmatrix}\right\}$$

$$=\begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \cdots & \cdots & \cdots \\ \sigma_{m1} & \cdots & \sigma_{mn} \end{bmatrix}$$

$$\text{where } \sigma_{ij}=\operatorname{cov}(X_i,Y_j)$$

Basic results

- If **A, B** are constant matrics and **Y** is a random vector,

$$\underset{k\times 1}{\mathbf{W}} = \mathbf{AY} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{m\times 1}{\mathbf{V}} = \mathbf{BY} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$E\left\{\mathbf{W}\right\} = \mathbf{A}E\left\{\mathbf{Y}\right\} \qquad E\left\{\mathbf{V}\right\} = \mathbf{B}E\left\{\mathbf{Y}\right\}$$

$$\operatorname{var}\left\{\mathbf{W}\right\} = E\left\{\left[\mathbf{AY} - \mathbf{A}E\left\{\mathbf{Y}\right\}\right]\left[\mathbf{AY} - \mathbf{A}E\left\{\mathbf{Y}\right\}\right]'\right\}$$

$$= E\left\{\left[\mathbf{A}\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)\right]\left[\mathbf{A}\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)\right]'\right\}$$

$$= E\left\{\left[\mathbf{A}\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)\right]\left[\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)'\mathbf{A}'\right]\right\}$$

$$= \mathbf{A}E\left\{\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)\left(\mathbf{Y} - E\left\{\mathbf{Y}\right\}\right)'\right\}\mathbf{A}' = \mathbf{A}\operatorname{var}\left\{\mathbf{Y}\right\}\mathbf{A}'$$

$$\operatorname{cov}\left\{\mathbf{W},\mathbf{V}\right\} = \operatorname{cov}\left\{\mathbf{AY},\mathbf{BY}\right\} = \mathbf{A}\operatorname{var}\left\{\mathbf{Y}\right\}\mathbf{B}'$$

Basic results

- If **A** is constant square matrix and **Y** is a random vector, and $E\{\mathbf{Y}\}=\boldsymbol{\mu}$, $\text{var}\{\mathbf{Y}\}=\boldsymbol{\Sigma}=\left(\sigma_{ij}\right)_{n\times n}$, then

$$E\{\mathbf{Y'AY}\}=\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}+tr(\mathbf{A}\boldsymbol{\Sigma})$$

Proof: $E\{\mathbf{Y'AY}\}=E\left\{\sum_{i,j=1}^{n}a_{ij}Y_iY_j\right\}=\sum_{i,j=1}^{n}a_{ij}E\{Y_iY_j\}$

$$=\sum_{i,j=1}^{n}a_{ij}\left(\mu_i\mu_j+\sigma_{ji}\right)=\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}+\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$$

Multivariate Normal Distribution

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

Multivariate Normal Density function: $\mathbf{Y} \sim N\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$,

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

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

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

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

Matrix Simple Linear Regression

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

with ε_i are i.i.d and $\varepsilon_i \sim N(0, \sigma^2)$.

Design matrix

- If we identify the following matrices,

$$\underset{n\times1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \underset{2\times1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \underset{n\times1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

$$\underset{n\times1}{\mathbf{Y}} = \underset{n\times2}{\mathbf{X}}\,\underset{2\times1}{\boldsymbol{\beta}} + \underset{n\times1}{\boldsymbol{\varepsilon}}$$

Matrix Normal Regression Model

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\,\underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} \sim N\left(\underset{n\times 1}{\mathbf{0}}, \sigma^2 \underset{n\times n}{\mathbf{I}}\right)$$

$$\underset{n\times 1}{E\left\{\mathbf{Y}\right\}} = \underset{n\times 2}{\mathbf{X}}\,\underset{2\times 1}{\boldsymbol{\beta}} + E\left\{\boldsymbol{\varepsilon}\right\} = \underset{n\times 2}{\mathbf{X}}\,\underset{2\times 1}{\boldsymbol{\beta}}$$

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

$$\mathbf{Y} \sim N\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

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

Regression Matrices

Matrices used in simple linear regression:

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

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

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

$$\text{Note: } \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + n\overline{X}^2, \quad \mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix}$$

$$\Rightarrow \quad \left|\mathbf{X'X}\right| = n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 = n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = nSS_{XX}$$

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

Finding the inverse of matrix A in R: solve(A)

Estimating Parameters by Least Squares

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2, \quad (b_0, b_1) = \arg\min Q(\beta_0, \beta_1)$$

Normal equations obtained from: $\dfrac{\partial Q}{\partial \beta_0} = 0, \dfrac{\partial Q}{\partial \beta_1} = 0,$

$$nb_0 + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i; \quad b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

$$\mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix}; \quad \mathbf{X'Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}; \quad \text{Let } \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$\Rightarrow \quad \mathbf{X'Xb} = \mathbf{X'Y} \quad \Rightarrow \quad \mathbf{b} = \left(\mathbf{X'X}\right)^{-1} \mathbf{X'Y}$$

Estimating Parameters by Least Squares

- Matrix derivation

$$\frac{\partial(\mathbf{A\beta})}{\partial\boldsymbol{\beta}}=\mathbf{A} \qquad \frac{\partial(\boldsymbol{\beta}'\mathbf{A})}{\partial\boldsymbol{\beta}}=\mathbf{A}'$$

$$\frac{\partial(\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\boldsymbol{\beta}'(\mathbf{A}+\mathbf{A}')$$

$$\mathbf{X}=(X_1,\cdots,X_m)', \quad \mathbf{Y}=(Y_1,\cdots,Y_k)'$$

$$\frac{\partial(\mathbf{Y})}{\partial\mathbf{X}}\hat{=}\begin{pmatrix}\frac{\partial Y_1}{\partial X_1} & \cdots & \frac{\partial Y_1}{\partial X_m}\\ \cdots & \cdots & \cdots\\ \frac{\partial Y_k}{\partial X_1} & \cdots & \frac{\partial Y_k}{\partial X_m}\end{pmatrix}$$

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

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

$$\frac{\partial Q}{\partial\boldsymbol{\beta}}=\left[\frac{\partial Q}{\partial\beta_0},\ \frac{\partial Q}{\partial\beta_1}\right]=-2\mathbf{Y'X}+2\boldsymbol{\beta}'\mathbf{X'X}$$

$$\frac{\partial Q}{\partial\boldsymbol{\beta}}=0 \quad\Rightarrow\quad \mathbf{X'Xb}=\mathbf{X'Y}$$

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

Distribution of estimator

$$\mathbf{b} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} \qquad \mathbf{Y} \sim N\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

$$E\{\mathbf{b}\} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'}E\{\mathbf{Y}\} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

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

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

$$\mathbf{b} \sim N\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X'X}\right)^{-1}\right)$$

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

Fitted value

$$\hat{Y}_i = b_0 + b_1 X_i, \;\; i = 1, 2, \ldots n$$

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix} = \mathbf{Xb} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} = \mathbf{HY}$$

Hat matrix: $\mathbf{H} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$

$$\boxed{h_{ij} = \frac{1}{n} + \frac{\left(X_i - \bar{X}\right)\left(X_j - \bar{X}\right)}{SS_{XX}}}$$

$$E\left\{\hat{\mathbf{Y}}\right\} = E\left\{\mathbf{HY}\right\} = \mathbf{H}E\left\{\mathbf{Y}\right\} = \mathbf{HX\beta} = \mathbf{X\beta}$$

$$\mathrm{var}\left\{\hat{\mathbf{Y}}\right\} = \mathrm{var}\left\{\mathbf{HY}\right\} = \mathbf{H}\,\mathrm{var}\left\{\mathbf{Y}\right\}\mathbf{H'} = \sigma^2\mathbf{H}$$

$$\hat{\mathbf{Y}} = \mathbf{HY} \sim N\left(\mathbf{X\beta}, \sigma^2\mathbf{H}\right)$$

Yhat之间并不独立

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

Property of hat matrix $\mathbf{H}$:

- $\mathbf{HY}=\hat{\mathbf{Y}}$; $\mathbf{HX}=\mathbf{X}$; $\mathbf{H}\hat{\mathbf{Y}}=\hat{\mathbf{Y}}$; $\mathbf{He}=\mathbf{0}$

$$\boxed{\begin{aligned}&\mathbf{HX}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}=\mathbf{X};\ \mathbf{H}\hat{\mathbf{Y}}=\mathbf{HXb}=\mathbf{Xb}=\hat{\mathbf{Y}}\\&\mathbf{H}\underset{n\times 1}{\mathbf{e}}=\mathbf{H}\left(\mathbf{Y}-\hat{\mathbf{Y}}\right)=\mathbf{HY}-\mathbf{H}\hat{\mathbf{Y}}=\hat{\mathbf{Y}}-\hat{\mathbf{Y}}=0\end{aligned}}$$

- symmetric

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

- idempotent

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

Residuals

$$e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \ldots n$$

$$\mathbf{e} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

- The matrix $\mathbf{I}-\mathbf{H}$ is also symmetric and idempotent.
  - $[\mathbf{I}-\mathbf{H}][\mathbf{I}-\mathbf{H}] = \mathbf{I}-\mathbf{H}-\mathbf{H}+\mathbf{H}^2 = \mathbf{I}-\mathbf{H}$
- The distribution of $\mathbf{e}$: $\mathbf{e} \sim N\left(\mathbf{0}, (\mathbf{I}-\mathbf{H})\sigma^2\right)$

$$E\{\mathbf{e}\} = E\{(\mathbf{I}-\mathbf{H})\mathbf{Y}\} = (\mathbf{I}-\mathbf{H})E\{\mathbf{Y}\} = (\mathbf{I}-\mathbf{H})\mathbf{X\beta} = \mathbf{X\beta} - \mathbf{X\beta} = \mathbf{0}$$

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

Analysis of Variance

Note: $\mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{n} Y_i^2$, $\mathbf{Y}'\mathbf{1} = \mathbf{1}'\mathbf{Y} = \sum_{i=1}^{n} Y_i$, $\left(\sum_{i=1}^{n} Y_i\right)^2 = (\mathbf{Y}'\mathbf{1})(\mathbf{1}'\mathbf{Y}) = \mathbf{Y}'\mathbf{J}\mathbf{Y}$, $\underset{n\times n}{\mathbf{J}} = \mathbf{1}\mathbf{1}'$

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

$$\Rightarrow SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$$

$$SSE = \mathbf{e}'\mathbf{e} = \left[(\mathbf{I}-\mathbf{H})\mathbf{Y}\right]'\left[(\mathbf{I}-\mathbf{H})\mathbf{Y}\right] = \mathbf{Y}'(\mathbf{I}-\mathbf{H})^2\mathbf{Y} = \mathbf{Y}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y}$$

$$SSR = SSTO - SSE = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} - \mathbf{Y}'\left[\mathbf{I}-\mathbf{H}\right]\mathbf{Y} = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$$

Analysis of Variance

- Quadratic forms for ANOVA

$$SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad \text{rank}\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right] = n-1$$

$$SSE = \mathbf{Y}'[\mathbf{I} - \mathbf{H}]\mathbf{Y} \qquad \text{rank}[\mathbf{I} - \mathbf{H}] = n-2$$

$$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad \text{rank}\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right] = 1$$

- $\mathbf{H}$, $\mathbf{J}/n$, $\mathbf{I}-\mathbf{J}/n$, $\mathbf{I}-\mathbf{H}$, $\mathbf{H}-\mathbf{J}/n$ are idempotent and symmetric

$$\boxed{\text{rank}[\mathbf{H}] = \text{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] = \text{tr}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\right] = \text{tr}[\mathbf{I}_2] = 2}$$

ANOVA in regression

- $\mathbf{I}-\mathbf{J}/n$, $\mathbf{I}-\mathbf{H}$, $\mathbf{H}-\mathbf{J}/n$ are symmetric and idempotent.

$$\text{rank}\left[\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right]=n-1 \qquad \text{rank}\left[\mathbf{I}-\mathbf{H}\right]=n-2 \qquad \text{rank}\left[\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right]=1$$

- Quadratic forms for ANOVA

$$SSTO=\mathbf{Y}'\left[\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}\sim\sigma^2\chi^2(n-1,\ \delta)$$

$$SSE=\mathbf{Y}'\left[\mathbf{I}-\mathbf{H}\right]\mathbf{Y}\sim\sigma^2\chi^2(n-2,\ 0)$$

$$SSR=\mathbf{Y}'\left[\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}\sim\sigma^2\chi^2(1,\ \delta)$$

$$\text{where } \delta=\frac{1}{\sigma^2}(\mathbf{X}\boldsymbol{\beta})'\left[\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{X}\boldsymbol{\beta}=\frac{\beta_1^2}{\sigma^2/SS_{XX}}.$$

Proof: Distribution of SSE

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} \sim N\left(\underset{n\times 1}{\mathbf{0}}, \sigma^2 \underset{n\times n}{\mathbf{I}}\right)$$

For simple regression, p=2

$$\underset{n\times 1}{\mathbf{e}} = (\mathbf{I}-\mathbf{H})\mathbf{Y} \sim N\left(\mathbf{0}, (\mathbf{I}-\mathbf{H})\sigma^2\right),$$

$$\exists \underset{n\times n}{\mathbf{U}} = \left(\underset{n\times(n-p)}{\mathbf{A}} \quad \underset{n\times p}{\mathbf{B}}\right), \quad \mathbf{U}'\mathbf{U} = \mathbf{I}_n$$

$$\mathbf{U}'(\mathbf{I}-\mathbf{H})\mathbf{U} = \operatorname{diag}\left(\underbrace{1,\cdots,1}_{\text{Number of 1}=n-p},0,\cdots 0\right)$$

$$\Rightarrow \mathbf{A}'\mathbf{A} = \mathbf{A}'(\mathbf{I}-\mathbf{H})\mathbf{A} = \mathbf{I}_{n-p}, \quad \mathbf{A}\mathbf{A}' = \mathbf{I}_n$$

$$\text{Let } \underset{(n-p)\times 1}{\mathbf{e}^*} = \mathbf{A}'\mathbf{e} \sim N\left(\mathbf{0}, \mathbf{I}_{n-p}\sigma^2\right) \Rightarrow \mathbf{e} = \mathbf{A}\mathbf{e}^*$$

$$SSE = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = \mathbf{e}^{*\prime}\mathbf{A}'\mathbf{A}\mathbf{e}^* = \mathbf{e}^{*\prime}\mathbf{e}^* = \sum_{i=1}^{n-p} e_i^{*2} \sim \sigma^2 \chi^2(n-p)$$

Inference in Regression Analysis

- Parameters

$$\mathbf{b} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} \sim N\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X'X}\right)^{-1}\right)$$

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

- Estimated mean response at $X = X_h$

$$\hat{Y}_h = b_0 + b_1 X_h = \mathbf{X_h b} \qquad \mathbf{X_h} = \begin{bmatrix} 1 & X_h \end{bmatrix}$$

$$\boldsymbol{\sigma}^2\left\{\hat{Y}_h\right\} = \mathbf{X_h}\boldsymbol{\sigma}^2\{\mathbf{b}\}\mathbf{X}'_\mathbf{h} = \sigma^2\left(\mathbf{X_h}\left(\mathbf{X'X}\right)^{-1}\mathbf{X}'_\mathbf{h}\right)$$

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

Predicted New Response at $X = X_h$:

$$\hat{Y}_h = b_0 + b_1 X_h = \mathbf{X_h b} \qquad s^2\{\text{pred}\} = MSE\left(1 + \mathbf{X_h}\left(\mathbf{X'X}\right)^{-1}\mathbf{X}'_\mathbf{h}\right)$$

Appendix: Cochran's Theorem

(Cochran's Theorem) Let $X_1, X_2, \ldots, X_n$ be independent $N(\mu_i, \sigma^2)$,.i.e $\mathbf{X} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, and

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \cdots + Q_k,$$

where $Q_i = \mathbf{X}'\mathbf{A}_i\mathbf{X}$, $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are symmetric and **idempotent** $n \times n$ matrices with rank$(\mathbf{A}_i) = r_i$, $i=1,2,\ldots,k$.

Then: (1) $Q_1, Q_2, \ldots, Q_k$ are independent.

$$(2)\quad \frac{Q_i}{\sigma^2} \sim \chi^2(r_i,\ \delta_i) \quad \text{with } \delta_i = \boldsymbol{\mu}'\mathbf{A}_i\boldsymbol{\mu}' / \sigma^2.$$

Corollary: Let $\mathbf{X} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, $\mathbf{A}$ is symmetric with rank$(\mathbf{A}) = r$, and $\delta = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}' / \sigma^2$. Then

$$\mathbf{X}'\mathbf{A}\mathbf{X} / \sigma^2 \sim \chi^2(r, \delta) \quad \Leftrightarrow \quad \mathbf{A} \text{ is } \textbf{idempotent}$$

Homework

- Page 210~212:

5.5, 5.18, 5.24, 5.28