# Ch3 Multiple regression

$$e_i \overset{iid}{\sim} N(0,\sigma^2)$$

simple: $Y_i = \beta_0 + \beta_1 X_i + e_i \qquad i = 1, 2 \cdots n$

multiple: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + e_i$

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_p X_{p1} + e_1 \\
Y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} \cdots + e_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} \cdots \beta_p X_{pn} + e_n
\end{aligned}$$

Matrix representation

$$\underset{n\times 1}{Y} = \underset{n\times(p+1)}{X}\,\underset{(p+1)\times 1}{\beta} + \underset{n\times 1}{e}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{p1} \\ 1 & X_{12} & & \vdots \\ 1 & \vdots & & \vdots \\ 1 & X_{1n} & & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

# How to define E & Var for <u>Vector</u>

$$E(e) = E\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} E(e_1) \\ E(e_2) \\ \vdots \\ E(e_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

$$Var(e) = \begin{pmatrix} \overset{Var(e_1)}{Cov(e_1,e_1)} & Cov(e_1,e_2) & \cdots & Cov(e_1,e_n) \\ Cov(e_2,e_1) & \overset{Var(e_2)}{Cov(e_2,e_2)} & & \\ Cov(e_3,e_1) & & \ddots & \\ \vdots & & & \\ Cov(e_n,e_1) & & & \overset{Var(e_n)}{Cov(e_n,e_n)} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots \\ 0 & \sigma^2 & & \mathrm{O} \\ & \mathrm{O} & \ddots & \\ & & & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

identity matrix size n

# Matrix Notation for Multiple Regression

- multiple linear regression in matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

- Least Square estimation for β

$$RSS(\beta) = \sum (y_i - \hat{y}_i)^2$$

$\mathbf{X}\beta$

$a = (a_1, \cdots a_n)$

$a' a = \sum_{i=1} a_i^2$

$$\stackrel{\star}{=} \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \quad \vdots \\ y_n - \hat{y}_n \end{pmatrix}' \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \quad \vdots \\ y_n - \hat{y}_n \end{pmatrix}$$

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

$\rightarrow Y'Y - Y'X\beta - \beta' X' Y + \beta' X' X\beta$

They are scalar

$$= \mathbf{Y}'\mathbf{Y} - 2\,\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \qquad (\mathbf{Y}'\mathbf{X}\beta = (\mathbf{Y}'\mathbf{X}\beta)' = \beta'\mathbf{X}'\mathbf{Y})$$

# Matrix Notation for Multiple Regression

- multiple linear regression in matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

- Least Square estimation for β

$$RSS(\beta) = \sum (y_i - \hat{y}_i)^2$$

$$= \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \quad \vdots \\ y_n - \hat{y}_n \end{pmatrix}' \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \quad \vdots \\ y_n - \hat{y}_n \end{pmatrix}$$

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

They are scalar

$$= \mathbf{Y}'\mathbf{Y} - \underbrace{2\,\mathbf{Y}'\mathbf{X}\beta}_{V\beta} + \underbrace{\beta'\mathbf{X}'\mathbf{X}\beta}_{\beta' M \beta} \qquad (\mathbf{Y}'\mathbf{X}\beta = (\mathbf{Y}'\mathbf{X}\beta)' = \beta'\mathbf{X}'\mathbf{Y})$$

# Matrix Differentiation

- Remember
  - $$\frac{\partial}{\partial \beta} c' \beta = c$$
  - $$\frac{\partial}{\partial \beta} \beta' c = c$$

- e.g.5
  - **β** =[$\beta_1$, $\beta_2$, $\beta_3$]
  - f(**β**)= **β'M β**
  - By Product Rule,

$$\underbrace{\sum_{j}^{n}\sum_{i}^{n}} \beta_i M_{ij} \beta_j$$

$$\frac{\partial f(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \beta' \mathrm{M} \beta$$

$$= (\beta' \mathrm{M})' + \mathrm{M} \beta$$

$$= (\mathrm{M}' + \mathrm{M}) \beta$$

**Key Results**

$$\boxed{\frac{\partial}{\partial \beta} c' \beta = \frac{\partial}{\partial \beta} \beta' c = c}$$

$$\boxed{\frac{\partial}{\partial \beta} \beta' \mathrm{M} \beta = (\mathrm{M}' + \mathrm{M}) \beta}$$

$$\frac{d\,ax}{dx} = a \qquad \frac{d}{dx} ax^2 = 2ax$$

# Least Square estimator

- Least Square Estimator:
  - Minimizes

$$RSS(\beta) = (Y - X\beta)'(Y - X\beta) = Y'Y - 2Y'X\beta + \beta'X'X\beta$$

  - Find Minimum by differentiation

$$\frac{\partial}{\partial \beta} c'\beta = c \qquad \frac{\partial}{\partial \beta} \beta' \mathrm{M} \beta = (\mathrm{M}' + \mathrm{M})\beta$$

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2(Y'X)' + (X'X + (X'X)')\beta$$

$$= -2X'Y + 2X'X\beta$$

$$\hat{\beta}_1 = \sum \underbrace{a_i}_{X \text{ only}} Y_i$$

  - Set the derivative equal 0 gives

$$\boxed{\hat{\beta} = \underbrace{(X'X)^{-1}X'}_{X \text{ only}}Y} = AY$$

# Probability Calculation of Matrix

- Mean
  - $$E(AX) = E\begin{pmatrix} \sum a_{1i}x_i \\ \sum a_{2i}x_i \\ \cdots \\ \sum a_{mi}x_i \end{pmatrix} = \begin{pmatrix} \sum a_{1i}E(x_i) \\ \sum a_{2i}E(x_i) \\ \cdots \\ \sum a_{mi}E(x_i) \end{pmatrix} = AE(X)$$

**A is mxm constant matrix**
**X is mx1 random vector**

- Variance
  - $$\begin{aligned} Var(AX) &= E([AX - E(AX)][AX - E(AX)]') \\ &= E([A(X - E(X))][A(X - E(X))]') \\ &= AE[(X - E(X))(X - E(X))']A' \\ &= AVar(X)A' \end{aligned}$$

$$\boxed{E(AX) = AE(X)}$$

$$\boxed{Var(AX) = AVar(X)A'} \neq A^2 Var(X)$$

$$E(aX) = a\,E(X)$$

$$Var(aX) = a^2\,Var(X)$$

# Properties Least Square estimator

- Model

$$Y = X\beta + e, \qquad \mathrm{E}(e) = 0, Var(e) = \sigma^2 I$$

- Least Square Estimator (LSE)

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

- Mean of LSE (Unbiasedness: mean of estimate= truth)

$$E(\hat{\beta}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'E(X\beta + e)$$

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

$Var(X\beta + e) = Var(e) = \sigma^2 I$

- Variance of LSE

$$Var(\hat{\beta}) = Var[\overbrace{(X'X)^{-1}X'}^{A}\overbrace{Y}^{Y}] = \overbrace{(X'X)^{-1}X'}^{A}\underset{\downarrow}{Var(Y)}\overbrace{X(X'X)^{-1}}^{A'}$$

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

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

$\left(Var(\hat{\beta}_1) = \frac{\sigma^2}{Sxx}\right)$

# A Matrix operator -- Trace

- Trace (tr) is the sum of diagonal element of a Square matrix
  - $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}$$ $$\boxed{tr(A) = \sum_{i=1}^{m} a_{ii}}$$
- Properties
  - $$tr(A+B) = \sum_{i=1}^{m} a_{ii} + b_{ii} = tr(A) + tr(B)$$
  - $$tr(AB) = \sum_{i=1}^{m} \underbrace{\sum_{j=1}^{n} a_{ij} b_{ji}}_{\text{Diagonal of AB}} = \sum_{j=1}^{n} \underbrace{\sum_{i=1}^{m} b_{ji} a_{ij}}_{\text{Diagonal of BA}} = tr(BA)$$
  - $$tr(E(A)) = \sum E(a_{ii}) = E(\sum a_{ii}) = E(tr(A))$$

$AB \neq BA$

$t(AB) = tr(BA)$

# Properties Least Square estimator

$a'a = \sum a_i^2$

- Model $\quad Y = X\beta + e, \qquad \mathrm{E}(e) = 0, Var(e) = \sigma^2 I$
- Residual sum of square

$$RSS(\hat\beta) = (Y - X\hat\beta)'(Y - X\hat\beta) = Y'Y - 2Y'X\hat\beta + \hat\beta'X'X\hat\beta$$

$(X'X)^{-1}X'Y$

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

$$= Y'(\underbrace{I - X(X'X)^{-1}X'}_{A})Y$$

- Note that

$E(Y) = X\beta$

$$E(Y'AY) = E(tr(Y'AY)) = E(tr(AYY')) = tr(AE(YY'))$$

$E(X\beta e') = X\beta E(e') = 0$

$$= tr(AE[(X\beta + e)(X\beta + e)']) = tr(A(X\beta\beta'X' + \sigma^2 I))$$

$Var(Y) = E(Y - E(Y))(Y - E(Y))'$

$= E(YY') + \cdots$

$E(ee') = Var(e)$

$$= tr(A(X\beta\beta'X')) + \sigma^2 tr(A)$$

- Put $A = I - H = I - X(X'X)^{-1}X'$
  - $\mathrm{tr}(AX\beta\beta'X) = (I_n - X(X'X)^{-1}X')X\beta\beta'X = (X - X)\beta\beta'X = 0$
  - $\mathrm{tr}(A) = tr(I_n - X(X'X)^{-1}X') = tr(I_n) - tr(X(X'X)^{-1}X')$

$$= tr(I_n) - tr((X'X)^{-1}X'X) = tr(I_n) - tr(I_{p+1}) = n - (p+1)$$

$X = \begin{pmatrix} 1 & x \cdots & x \\ \vdots & \vdots & \vdots \\ 1 & & \end{pmatrix}$

$n \times (p+1)$

$X'X \quad (p+1)\cdot(p+1)$

$$\boxed{E(RSS(\hat\beta)) = \sigma^2(n - (p+1)) \quad \Rightarrow \quad \hat\sigma^2 = \frac{RSS(\hat\beta)}{n - (p+1)}}$$

$Y'AY = (a)$

$tr(Y'AY) = a$

$$\underset{1\times n}{Y'}\ \underset{n\times 1}{Y}$$

$$\underset{1\times n}{Y'}\ \underset{n(p+1)}{X}\ \underset{(p+1)(p+1)}{(X'X)^{-1}}\ \underset{(p+1)\cdot n}{X'}\ \underset{n\times 1}{Y}$$

$$= Y'\big(I_n - \underset{n(p+1)\ \cdots\ (p+1)n}{X(X'X)^{-1}X'}\big)Y$$

$$X(X'X)^{-1}X' \neq (X'X)^{-1}X'X = I_{p+1}$$

$$\not\parallel \quad n\times n$$

★ $tr\big(X(X'X)^{-1}X'\big) = tr\{(X'X)^{-1}X'X\} = p+1$