STA302/1001 - Methods of Data Analysis I

(Week 04- Lecture B)

Wei (Becky) Lin

June 7-8, 2017

UNIVERSITY OF TORONTO

Midterm and A3

- Please hand in your medical note for missing your midterm in 7 days (started from MT date), fail to do so get zero as stated in the course syllabus.
- Midterm result should be available over this weekend.
- A3 guideline of the course project is available on portal, it is due 11pm, June 22.
- Final is available on June 9 by art and sci.

Last Lecture

- Variable transformations.
- More on logarithmic transformation.
- Box-Cox transformation.
- Interpretation of slope after transformation.
- Chapter 4: Simultaneous Inferences

Review: from p-value to conclusion

True model $\qquad$ Estimated model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i; \quad \hat{Y}_i = b_0 + b_1 X_i$$

$$H_0 : \beta_1 = 0; \qquad H_a : \beta_1 \neq 0$$

- test statistic:

$$t^* = \frac{b_1}{s(b_1)}|_{H_0} \sim t_{n-2} \qquad b_1 = \frac{S_{xy}}{S_{xx}}, \quad s(b_1) = \left(\frac{MSE}{S_{xx}}\right)^{\frac{1}{2}}$$

- Let the P-value of this test be P_t, assume significance level, $\alpha = 5\%$
  - Case 1: $P_t < 0.05$
    - We have weak/moderate/strong evidence to reject the null hypothesis.
    - The linear relationship beween Y and X is significant at 5% level.
    - We have weak/moderate/strong evidence of a linear relationship between Y and X.
  - Case 2: $P_t > 0.05$ $\qquad$ It doesn't imply that H_0 is true*.
    - We failed to reject the null. (or, there is not enough evidence to reject the null.)
    - We don't have evidence to indicate a linear relationship between Y and X.

Week 04 Lecture B: Learning objectives & Outcomes

- Review on matrices.
  - What is matrix?
  - Matrix: transpose, multiplication, addition, inverse
  - Special matrices: identity, diagonal, column vector, row vector.
  - Variance-covariance matrix of a random vector.
  - Matrix Differentiation.
- Simple Linear Regression Model in Matrix form.
  - SLR model in matrix form.
  - Least Squares Estimation of Regression Parameters in matrix form.
  - Fitted values and residuals in matrix form.
  - Inference in Regression Analysis
  - ANOVA in matrix form

Review on Matrices

Review on Matrices

- Before we take on multiple linear regression, we first look at simple regression from a matrix perspective.
  - It is easy to generalize it to the general analysis of linear models.

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

$$A = \begin{bmatrix} 10 & 5 & 8 \\ 4 & 12 & 2 \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$$

  - Matrices are often denoted by captial letters.
  - Dimension of a matrix is: $(\#$ rows $)\times(\#$ columns $)$
  - Elements of matrices are referenced by subscripts (i, j), representing row and column index. e.g. $a_{1,3} = 8, a_{2,2} = 12$

Review on Matrices

- Before we take on multiple linear regression, we first look at simple regression from a matrix perspective.
  - It is easy to generalize it to the general analysis of linear models.

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

$$A = \begin{bmatrix} 10 & 5 & 8 \\ 4 & 12 & 2 \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$$

  - Matrices are often denoted by captial letters.
  - Dimension of a matrix is: $(\#$ rows $)\times(\#$ columns $)$
  - Elements of matrices are referenced by subscripts (i, j), representing row and column index. e.g. $a_{1,3} = 8, a_{2,2} = 12$

Review on Matrices (cont.)

- In general, a matrix can be represented in full or abbreviated form

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} = A_{2\times 3} = [a_{i,j}], i = 1, 2; j = 1, 2, 3$$

- Matrix with 1 column called column vector.

$$A_{2\times 1} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \end{bmatrix}$$

- Matrix with 1 row called row vector.

$$A_{1\times 3} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \end{bmatrix}$$

Review on Matrices (cont.)

- Matrix transpose: flipped version of matrix.
  - Denoted by A', or A^T.

$$A_{2\times 3} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$$

$$B_{3\times 2} = A' = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{2,1} \\ a_{1,2} & a_{2,2} \\ a_{1,3} & a_{2,3} \end{bmatrix},$$

  - That is, $b_{i,j} = a_{j,i}, j = 1, 2; i = 1, 2, 3$
  - Transpose of a column vector is a row vector.

$$A_{2\times 1} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \end{bmatrix} \Rightarrow A' = \begin{bmatrix} a_{1,1} & a_{2,1} \end{bmatrix}$$

  - Matrix equality: two matrices are called equal if they have same dimensions and the corresponding elements are equal.
- $(ABC)^T = C^T B^T A^T$

Review on Matrices (cont.)

- Matrix addition: the sum of two matrices with equal dimensions is matrix of the sum of their corresponding elements.

$$A_{m\times n} + B_{m\times n} = C_{m\times n} \Leftrightarrow [c_{i,j}] = [a_{i,j} + b_{i,j}]$$

  - For example, sum of two column vectors is another column vector

$$A_{2\times 1} + B_{2\times 1} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

- Matrix multiplication: product of two matrices is matrix of cross-product of their corresponding rows and columns.

$$A_{m\times n} \cdot B_{n\times p} = C_{m\times p} \Leftrightarrow [c_{i,j}] = \left[\sum_{k=1}^{n} a_{i,k} b_{k,j}\right], \forall i = 1, \ldots, m, j = 1, \ldots, p$$

$$A_{1\times 2} B_{2\times 2} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 3 & 1\cdot 2 + 2\cdot 4 \end{bmatrix} = \begin{bmatrix} 7 & 10 \end{bmatrix}$$

Review on Matrices (cont.)

Matrix multiplication:

- # columns for A must match # rows for B.
- Matrix multiplication is NOT symmetric.

$$A_{1\times 2} \cdot B_{2\times 2} \neq B_{2\times 2} \cdot A_{1\times 2}$$

- Multiplication by a scalar (number)

$$c \cdot A_{2\times 2} = 3 \cdot \begin{bmatrix} 6 & 2 \\ 7 & 4 \end{bmatrix} = \begin{bmatrix} 3*6 & 3*2 \\ 3*7 & 3*4 \end{bmatrix} \begin{bmatrix} 18 & 6 \\ 21 & 12 \end{bmatrix}$$

Diagonal matrix:

- It is a square matrix and all off-diagonal elements are 0.

$$D_{3\times 3} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}, \qquad I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Identity matrix: a special case of diagonal matrix, with all diagonal elements equal to 1.

Identity Matrix

- Multiplying matrix **A** by identity matrix **I** leaves **A** unchanged.
- **IA=AI=A**

$$I \cdot A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 1 \\ 5 & 1 & 3 \\ 7 & 9 & 12 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 1 \\ 5 & 1 & 3 \\ 7 & 9 & 12 \end{bmatrix} = A \cdot I$$

- Scalar matrix: diagonal matrix with all diagonal elements being equal

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

Inverse of a Matrix

Assume A is invertible, that is, the inverse of **A** exists.

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

For a 2 by 2 matrix, a simple formula exists to find its inverse

$$\textit{if} \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

$$\textit{then} \quad A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- Note that the quantity $ad - bc$ is the determinant of A.
- $1/(ad - bc)$ is not defined when $ad - bc = 0$ since it is never possible to divide by zero. It is for this reason that the inverse of A does not exist if the determinant of A is zero.

Inverse of a Matrix (cont.)

- Not every square matrix **A** has an inverse.
  - Matrix must be non-singular (full rank).
  - Matrix rank is maximum number of linearly indepedent columns. (equivalently rows).

$a_3 = a_1 - 2a_2$

$$A_{3\times 3} = \begin{bmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 0 & 2 \\ 1 & 1 & -1 \end{bmatrix}$$

  This leads to rank(**A**)=2, since $a_3 = a_1 - 2a_2$.
  - Square matrix is full-rank if $rank(\mathbf{A}_{n\times n}) = n$.
- $(AB)^{-1} = B^{-1}A^{-1}$ assume the square matrices A and B, are both invertible.
- $\{(A)^{-1}\}' = \{A'\}^{-1}$

Column Vectors

- A column with all elements 1 and zero are **1** and **0**

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

- A square matrix with all elements 1: J

$$J_{r\times r} = \mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

- $\mathbf{1}'\mathbf{1} = n$

Mean of a random Vector

- Vecotr of random variables (RV's)

$$\mathbf{Y_{n\times 1}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

  where $Y_1, \ldots, Y_n$ are RV's.
- Mean of random vector: $E(Y_i) = \mu_i, i = 1, \ldots n$

$$E(\mathbf{Y})_{\mathbf{n\times 1}} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \mu_{\mathbf{n\times 1}}$$

Mean of a linear combination of Random Vector

- Linear combination of random vectors
  - Constant matrix, vector: $\mathbf{A}_{r\times n}, b_{r\times 1}$
  - Random vector: $\mathbf{Y}_{n\times 1}$ with $E(\mathbf{Y}) = \mu_y, Var(\mathbf{Y}) = \Sigma_Y$.

$$\underbrace{\mathbf{X_{r\times 1}}}_{\text{RV}} = \mathbf{b_{r\times 1}} + \mathbf{A_{r\times n}}\,\underbrace{\mathbf{Y_{n\times 1}}}_{\text{RV}}$$

some constant

- For example

$$\mathbf{b_{2\times 1}} + \mathbf{A_{2\times 3}Y_{3\times 1}} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$\mathbf{X_{2\times 1}} = \begin{bmatrix} b_1 + a_{1,1}Y_1 + a_{1,2}Y_2 + a_{1,3}Y_3 \\ b_2 + a_{2,1}Y_1 + a_{2,2}Y_2 + a_{2,3}Y_3 \end{bmatrix}$$

- Mean of $\mathbf{X}$: $\boldsymbol{\mu_x} = E(\mathbf{X}) = E(\mathbf{b+AY}) = \mathbf{b} + \mathbf{A}E(\mathbf{Y}) = \mathbf{b} + \mathbf{A}\boldsymbol{\mu_y}$

Var-Cov of Random Vectors

- Variance-covariance matrix of random vector

$$Var(\mathbf{Y}) = Var(\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{n\times 1}) = \begin{bmatrix} V(Y_1) & Cov(Y_1, Y_2) & \dots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & V(Y_2) & \dots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \dots & Var(Y_n) \end{bmatrix}_{n\times n} = \Sigma_Y$$

$\Sigma_Y :=$ var-cov mat of $\vec{Y}$

① on diagonal : $Var(Y_i)$

② off diagonal : $Cov(Y_i, Y_j)$ $i \neq j$

③ $\Sigma_Y^T = \Sigma_Y$

- Variance-covariance matrix is symmetric
  - $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$, that is $\Sigma = \Sigma^T$
- variance-covariance for $\boldsymbol{\epsilon}$

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

① $\sigma^2 = V(Y_1) = \cdots = V(Y_n)$
constant variance

② $Cov(Y_i, Y_j)_{i \neq j} = 0$: errors are uncorrelated

Var-Cov of Random Vectors

Show $Var(\mathbf{Y}) = E\{[\mathbf{Y}-\mu]\cdot[\mathbf{Y}-\mu]'\}$

$$Var(\mathbf{Y}) = \Sigma_Y = E\{[\mathbf{Y}-\mu]\cdot[\mathbf{Y}-\mu]'\}$$

$$= E\left(\begin{bmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \\ \vdots \\ Y_n-\mu_n \end{bmatrix}_{n\times 1} \begin{bmatrix} Y_1-\mu_1 & Y_2-\mu_2 & \dots & Y_n-\mu_n \end{bmatrix}_{1\times n}\right) \Rightarrow n\times n \text{ matrix}$$

$$= E\left(\begin{bmatrix} (Y_1-\mu_1)^2 & (Y_1-\mu_1)(Y_2-\mu_2) & \dots & (Y_1-\mu_1)(Y_n-\mu_n) \\ \vdots & \vdots & \vdots & \vdots \\ (Y_n-\mu_n)(Y_1-\mu_1) & (Y_n-\mu_n)(Y_2-\mu_2) & \vdots & (Y_n-\mu_n)^2 \end{bmatrix}\right)$$

$$= \begin{bmatrix} V(Y_1) & Cov(Y_1,Y_2) & \dots & Cov(Y_1,Y_n) \\ Cov(Y_2,Y_1) & V(Y_2) & \dots & Cov(Y_2,Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n,Y_1) & Cov(Y_n,Y_2) & \dots & Var(Y_n) \end{bmatrix}$$

$$= \Sigma_{n\times n}$$

Var-Cov Random Vectors

- Variance-covariance matrix for uncorrelated or independent vector of RV's
  - $Y_1, \ldots, Y_n$ with $Cov(Y_i, Y_j) = 0, \forall i \neq j$.

$$Var(\mathbf{Y}) = \Sigma_Y = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{bmatrix}$$

  - It is a diagonal matrix: all off-diagonal elements are 0.
- W, Y: random vectors
- A: a constant matrix
- We then have
  - E(A)=A
  - W=AY, E(W)=E(AY)=AE(Y)
  - W=AY, Var(W)=Var(AY)=$AV(Y)A^T$ *

Example: Var-Cov of linear combination of random vector

- Let $\mathbf{X} = \mathbf{b} + \mathbf{AY}$.
- Using $(A+B)C = A\overset{C}{\cancel{B}} + \overset{B}{\cancel{A}}C$, $(AB)^T = B^T A^T$,
- Show the var-cov matrix of X is

$$\Sigma_X = E\{(\mathbf{X} - \mu_{\mathbf{x}})(\mathbf{X} - \mu_{\mathbf{x}})^T\} = A\Sigma_y A^T \qquad (*)$$

proof: · $X = b + AY$

· $E(X) = E(b + AY) = b + AE(Y)$

· $X - E(X) = (b + AY) - (b + AE(Y)) = AY - AE(Y) = A(Y - \mu_Y)$

Now by def'n

$$\Sigma_X = E((X-\mu_X)(X-\mu_X)^T) = E\{A(Y-\mu_Y)\underbrace{[A(Y-\mu_Y)]^T}\}$$

$$= E\{A(Y-\mu_Y)\underbrace{(Y-\mu_Y)^T A^T}\} \quad \because (AB)^T = B^T A^T$$

$$= A\, E\{(Y-\mu_Y)(Y-\mu_Y)^T\} A^T$$

$$= A\Sigma_Y A^T$$

Q.E.D.

Example 2 on Var-Cov

$\Rightarrow$ imply $\begin{cases} \cdot\ Y_1 \sim N(0, \sigma_1^2) \\ \cdot\ Y_2 \sim N(0, \sigma_2^2) \\ \cdot\ cov(Y_1, Y_2) = \sigma_{1,2} \end{cases}$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{bmatrix}), \quad \mathbf{Y} \sim N(\mathbf{0}, \mathbf{\Sigma})$$

Show $V(a_1 Y_1 + a_2 Y_2) = a_1^2 \sigma^2 + a_2 \sigma_2^2 + 2a_1 a_2 \sigma_{1,2}$

Proof: Method 1:

- $V(a_1Y_1 + a_2Y_2) = V(a_1Y_1) + V(a_2Y_2) + 2\,cov(a_1Y_1, a_2Y_2)$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\,cov(Y_1, Y_2)$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\,\sigma_{1,2}$$

Method 2:

- $W = a_1Y_1 + a_2Y_2 = (a_1 \ \ a_2)\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = A_{1\times 2} Y_{2\times 1}$

$\Rightarrow Var(W) = Var(AY) = A\Sigma_Y A^T$ by (*) - slide 21

$$= (a_1 \ \ a_2)\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\,\sigma_{1,2}$$

Q.E.D.

Example 3 on Var-Cov

- Example: $Y_1, Y_2 \sim N(0,1)$ independently.
- Show

$$\mathbf{W} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}}_{Y} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix})$$

Proof: $W = AY$, and $\mu_Y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. $\Sigma_Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$

- $E(W) = E(AY) = A\,E(Y) = 0_{2\times 1}$
- $Var(W) = Var(AY) = A \Sigma_Y A^T$

$$= A I A^T = AA^T$$

$$= \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$ Q.E.D.

Matrix Differentiation

$$\underset{m\times 1}{Y} = f(X), \text{where } x_{n\times 1}$$

Then the $m \times n$ matrix of first-order partial derivatives of the transformation from x to Y:

$$\boxed{\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}} = \begin{bmatrix} \frac{\partial Y_1}{\partial \mathbf{X}} \\ \frac{\partial Y_2}{\partial \mathbf{X}} \\ \vdots \\ \frac{\partial Y_n}{\partial \mathbf{X}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}_{m\times n}$$

Matrix Differentiation (cont.)

$$MD01: \underset{m\times 1}{Y} = \underset{m\times n}{A} \ \underset{n\times 1}{X} \Rightarrow \frac{\partial Y}{\partial X} = A$$

Proof:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = AX$$

$$\Rightarrow Y_i = \sum_{k=1}^{n} a_{ik} X_k$$

$$\Rightarrow \frac{\partial Y_i}{\partial X_j} = a_{ij} \quad \forall i = 1, 2, \cdots m \text{ and } j = 1, 2, \cdots n$$

$$\Rightarrow \frac{\partial Y_i}{\partial X} = [a_{i1}, a_{i2}, \cdots a_{in}]$$

$$\Rightarrow \boxed{\frac{\partial Y}{\partial X} = A}$$

Q.E.D.

Matrix Differentiation (cont.)

$$MD02: \underset{1\times1}{\alpha} = \underset{1\times m}{Y^T} \underset{m\times n}{A} \underset{n\times1}{X} \Rightarrow \boxed{\frac{\partial\alpha}{\partial X} = Y^T A,} \text{ and } \boxed{\frac{\partial\alpha}{\partial Y} = X^T A^T}$$

Proof:

① $W^T = Y^T A \Rightarrow \alpha = W^T X$

$\Rightarrow \frac{\partial \alpha}{\partial X} = W^T$ by MD01

$= Y^T A$

② $\alpha_{1\times1} \Rightarrow \alpha = \alpha^T = X^T A^T Y$

$\Rightarrow \frac{\partial \alpha}{\partial Y} = X^T A^T$

Q.E.D.

Matrix Differentiation (cont.)

For the special case in which the scalar α is given by the quadratic form

$$MD03: \underset{1\times 1}{\alpha} = \underset{1\times n}{X^T} \; \underset{n\times n}{A} \; \underset{n\times 1}{X} \Rightarrow \frac{\partial \alpha}{\partial X} = X^T(A + A^T)$$

Proof:

by def'n $\alpha = (x_1, x_2, \cdots x_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

$$\Rightarrow \frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

$$= (x_1, x_2, \cdots x_n) \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{pmatrix} + (x_1, \cdots x_n) \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{nk} \end{pmatrix}$$

$$= X^T A^T_{\text{row } k} + X^T A_{\text{column } k}, \quad \text{for } k = 1, 2, \cdots n$$

$$\Rightarrow \boxed{\frac{\partial \alpha}{\partial X} = X^T A^T + X^T A = X^T (A^T + A)}$$

Matrix Differentiation (cont.)

For the special case in which the scalar α is given by the quadratic form, and the A is a symmetric matrix, $\Rightarrow A = A^T$

$$MD04: \underset{1\times 1}{\alpha} = \underset{1\times n}{X^T} \underset{n\times n}{A} \underset{n\times 1}{X} \Rightarrow \frac{\partial \alpha}{\partial X} = 2X^T A$$

Proof:

From MD03 we have

$$\frac{\partial \alpha}{\partial X} = X^T(A + A^T)$$

$$= X^T(A + A) \quad \text{since } A = A^T$$

$$= 2X^T A$$

Q.E.D.

Matrix approach to SLR

SLR in matrix form

- **Y**: consisting of the n observations on the response variables.

$$Y_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

- **X** Matrix:

$$X_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

often referred to as the design matrix.

SLR in matrix form

$X^T X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix}_{2\times n} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}_{n\times 2} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$

- Product

$= (Y_1, Y_2 \cdots Y_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sum_{i=1}^n Y_i^2$

$$\underset{1\times 1}{\mathbf{Y^T Y}} = \sum_i^n Y_i^2 \text{ : a scalar, a sum of square terms}$$

$$\underset{2\times 1}{\mathbf{X^T Y}} = \begin{bmatrix} \sum Y_i \\ \sum (X_i Y_i) \end{bmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix}_{2\times n} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} \sum_i^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}$$

$$\underset{1\times 1}{\mathbf{X^T X}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \Rightarrow \det(X^T X) = n\sum X_i^2 - (n\bar{X})^2$$

$$\Rightarrow \underset{1\times 1}{(\mathbf{X^T X})^{-1}} = \begin{bmatrix} \underbrace{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}}_{= a_{11}} & -\frac{\bar{X}}{\sum (X_i - \bar{X})^2} \\ -\frac{\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix}$$

$= n(\sum X_i^2 - n\bar{X}^2)$

$= n\sum_{i=1}^n (X_i - \bar{X})^2$

$= n S_{XX}$

$= \frac{1}{\det(X^T X)} \begin{pmatrix} \sum X_i^2 & -n\bar{X} \\ -n\bar{X} & n \end{pmatrix} =$

$a_{11} = \frac{\sum X_i^2}{n S_{XX}} = \frac{(\sum X_i^2 - n\bar{X}^2) + n\bar{X}^2}{n S_{XX}} = \frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}$

SLR in matrix form

- Simple linear regression: mean response in matrix form

$$E(Y_i) = \beta_0 + \beta_1 X_i, i = 1, \ldots, n$$

- Let

$$E(\mathbf{Y})_{\mathbf{n}\times\mathbf{1}} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}, \mathbf{X}_{\mathbf{n}\times\mathbf{2}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \boldsymbol{\beta}_{\mathbf{2}\times\mathbf{1}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- Columns represent variables.
- Rows represent observations.

SLR in matrix form (cont.)

- SLR: Mean response in matrix form

$$E(\mathbf{Y})_{\mathbf{n}\times\mathbf{1}} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \mathbf{X}_{\mathbf{n}\times\mathbf{2}} \cdot \boldsymbol{\beta}_{\mathbf{2}\times\mathbf{1}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

SLR model in matrix form (cont.)

- Simple linear regression in matrix form

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

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

That is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- For $\mathbf{Y}, \mathbf{X}$
  - Columns represent variables.
  - Rows represent observations.

SLR model in matrix form (cont.)

- The normal error regression model

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

  - $\boldsymbol{\epsilon}$: a vector of independent normal r.v.
  - $E(\boldsymbol{\epsilon}) = 0$
  - $V(\boldsymbol{\epsilon}) = \sigma^2 I$
  - Above 3 conditions on the error term can be written as

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) = N(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix})$$

or in a very brief way, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$

Least Squares Estimation of Regression Parameters

- Normal Equations

$$nb_0 + b_1 \sum X_i = \sum Y_i \tag{1}$$
$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i \tag{2}$$

This is equivalent in matrix form

$$\underset{2\times 2}{X'X}\,\underset{2\times 1}{b} = \underset{2\times 1}{X'Y}$$

$$X'X \cdot b = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

- The vector of the least squares regression coefficients

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & -\frac{\bar{X}}{\sum (X_i - \bar{X})^2} \\ -\frac{\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Verify the LS matrix form

Exercise: ① $\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) = \Sigma(X_iY_i) - n\bar{X}\bar{Y}$
② $\Sigma(X_i-\bar{X})^2 = \Sigma X_i^2 - n\bar{X}^2$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & -\frac{\bar{X}}{\sum(X_i-\bar{X})^2} \\ -\frac{\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Proof: Let $S_{XX} = \sum_{i=1}^n (X_i-\bar{X})^2$

- $S_{XY} = \Sigma(X_i-\bar{X})(Y_i-\bar{Y})$
$= \Sigma(X_iY_i) - n\bar{X}\bar{Y}$
- $b_1 = \frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} = \frac{S_{XY}}{S_{XX}}$

$$\begin{pmatrix} \frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} & -\frac{\bar{X}}{S_{XX}} \\ -\frac{\bar{X}}{S_{XX}} & \frac{1}{S_{XX}} \end{pmatrix} \begin{pmatrix} \sum_1^n Y_i \\ \sum_1^n X_iY_i \end{pmatrix}$$

$$= \begin{pmatrix} \frac{n\bar{Y}}{n} + \frac{n\bar{Y}\bar{X}^2}{S_{XX}} - \frac{\bar{X}\sum_{i=1}^n X_iY_i}{S_{XX}} \\ -\frac{\bar{X}n\bar{Y}}{S_{XX}} + \frac{\Sigma(X_iY_i)}{S_{XX}} \end{pmatrix} = \begin{pmatrix} \bar{Y} - \bar{X}\frac{\Sigma X_iY_i - n\bar{X}\bar{Y}}{S_{XX}} \\ \frac{\Sigma(X_iY_i) - n\bar{X}\bar{Y}}{S_{XX}} \end{pmatrix} \begin{matrix} = b_0 \\ = b_1 \end{matrix}$$

$$= \begin{pmatrix} \bar{Y} - b_1\bar{X} \\ b_1 \end{pmatrix} \qquad = \begin{pmatrix} \bar{Y} - b_1\bar{X} \\ S_{XY}/S_{XX} \end{pmatrix}$$

Least Squares Estimation of Regression Parameters

Minimize the quantity

- $X'X$ is symmetric

$\because (X'X)^T = X'(X')^T$
$= X'X$

$\begin{cases} Q: 1\times 1 \\ Y: n\times 1 \\ \beta: 2\times 1 \end{cases}$

$$
\begin{aligned}
Q &= \sum [Y_i - (\beta_0 + \beta_1 X_i)]^2 \\
&= (Y - X\beta)^T (Y - X\beta) \\
&= \underbrace{Y^T Y}_{1\times 1} - \underbrace{\beta^T}_{1\times 2}\underbrace{X^T}_{2\times n}\underbrace{Y}_{n\times 1} - \underbrace{Y^T}_{1\times n}\underbrace{X}_{n\times 2}\underbrace{\beta}_{2\times 1} + \overset{1\times 2}{\beta^T}\overset{2\times n}{X^T}\overset{n\times 2}{X}\overset{n\times 2}{\beta}
\end{aligned}
$$

① $\frac{\partial}{\partial\beta}(\beta^T X^T Y)$

$= \frac{\partial}{\partial\beta}(Y^T X\beta)$

$= Y'X$ (MD02)

To find the value of β that minimizes Q:

$$
\frac{\partial}{\partial\beta} Q = \left[\partial Q/\partial\beta_0 \quad \partial Q/\partial\beta_1\right] = -2Y'X + 2\beta^T X^T X
$$

② $\frac{\partial}{\partial\beta}(Y^T X\beta)$

$= Y'X$ (MD02)

- Setting this equation to zero: $-2Y'X + 2b^T X^T X = 0$
- Transpose both sides of the equation, we have

$$
X'Xb = X'Y
$$

③ $\frac{\partial}{\partial\beta}(\beta^T X^T X\beta)$

$= 2\beta^T(X^T X)$

(MD04)

- Solving it for $\hat{\beta}$. Assume $X'X$ is invertible

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

Fitted Values and Residuals

Fitted Values:

$$\underset{n\times 1}{\hat{\boldsymbol{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \underset{n\times 2}{\boldsymbol{X}}\underset{2\times 1}{\boldsymbol{b}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y}$$

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

Fitted Values and Residuals (conts.)

$$\hat{Y} = HY \iff \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \underbrace{\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix}}_{H} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},$$

$$h_{ij} = \frac{1}{n} + \frac{(X_i - \bar{X})(X_j - \bar{X})}{\sum (X_i - \bar{X})^2}$$

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}$$

Hat Matrix:

- $H = X(X'X)^{-1}X'$: Hat matrix or "projection" matrix.
- Important in diagnostics for regression
- Symmetric and idempotent:

$$H^T = H; \qquad HH = H$$

- $\sum_{j=1}^n h_{ij} = 1$
- $\sum_{j=1}^n h_{ij}^2 = h_{ii}$
- $h_{ij} = h_{ji}$
- $\frac{1}{n}\sum_{i=1}^n h_{ii} = \frac{2}{n}$

① $H^T = [X(X'X)^{-1}X']' = X[(X^TX)^{-1}]^T X^T$

$= X[(X^TX)^T]^{-1}X^T = X(X^TX)^{-1}X^T = H$

② $HH = X(X^TX)^{-1}X^T X(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$

$$\boxed{HH = H \iff ①\ \sum_{j=1}^n h_{ij}^2 = h_{ii} \quad ②\ \sum_{k=1}^n h_{ik}h_{kj} = h_{ij}}$$

Residuals

Residuals: $e_i = Y_i - \hat{Y}_i$

$$\underset{n\times 1}{\boldsymbol{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \underset{n\times 1}{\boldsymbol{Y}} - \underset{n\times 1}{\hat{\boldsymbol{Y}}} = \underset{n\times 1}{\boldsymbol{Y}} - \underset{n\times 1}{\boldsymbol{X}\boldsymbol{b}}$$

Variance-Covariance Matrix of Residuals:

- $\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$
- $(\boldsymbol{I} - \boldsymbol{H})$: symmetric and idempotent
- $\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{e}\}} = \sigma^2(\boldsymbol{I} - \boldsymbol{H}) \Rightarrow$ estimated by $\underset{n\times n}{\boldsymbol{s}^2\{\boldsymbol{e}\}} = MSE(\boldsymbol{I} - \boldsymbol{H})$

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

Residual distribution

$$\underset{n\times 1}{Y} \sim N(X\beta, \sigma^2 I), \Rightarrow \underset{n\times 1}{e} \sim N(0, \sigma^2(I-H))$$

Proof:

$e = Y - \hat{Y} = Y - HY = (I-H)Y$, where $H = X(X'X)^{-1}X'$

$\Rightarrow E((I-H)Y) = (I-H)E(Y)$

$= X\beta - \underbrace{H}X\beta = X\beta - \underbrace{\overbrace{X(X'X)^{-1}X'X}^{I}}\beta = X\beta - X\beta = 0$

$Var(e) = Var((I-H)Y)$

$= (I-H)\Sigma_Y(I-H)^T$

$= (I-H)\sigma^2 I(I-H)$

$= \sigma^2(I-H)(I-H) = \sigma^2(I-H)$

$I-H$ is idempotent

* For var-cov matrix: Symmetric.

Regression Coefficients:

- The variance-covariance matrix of **b**:

$$Var(b) = \begin{bmatrix} V(b_0) & V(b_0, b_1) \\ V(b_1, b_0) & V(b_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right) & -\sigma^2\frac{\bar{X}}{\sum(X_i-\bar{X})^2} \\ -\sigma^2\frac{\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{\sigma^2}{\sum(X_i-\bar{X})^2} \end{bmatrix}$$

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

$Var(b_0)$

$cov(b_0, b_1) = -\frac{\sigma^2\bar{X}}{S_{XX}}$

$Var(b_1) = \frac{\sigma^2}{S_{XX}}$

Inference in Regression Analysis (cont.)

- The estmiated variance-covariance of **b**:

$$
\begin{aligned}
\underset{2\times 2}{s^2(b)} &= MSE(X'X)^{-1} \\
&= MSE \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & -\frac{\bar{X}}{\sum(X_i-\bar{X})^2} \\ -\frac{\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix} \\
&= \hat{\sigma}^2(X'X)^{-1}
\end{aligned}
$$

- $b = (X'X)^{-1}X'Y = AY$
  - $\Rightarrow V(b) = V(AY) = A\Sigma_Y A^T = \sigma^2 AA^T = \sigma^2(X'X)^{-1}$

Mean response and prediction of new observation

- The mean response at X_h

$$\underset{2\times 1}{X_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix}; \qquad \hat{Y}_h = X_h' b$$

$$\begin{aligned} V(\hat{Y}_h) &= V(X_h' b) = X_h' V(b) X_h \\ &= X_h' \, \sigma^2 (X'X)^{-1} \, X_h \\ &= \sigma^2 X_h' (X'X)^{-1} X_h \\ &= \sigma^2 [\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}] \\ &\Rightarrow s^2(\hat{Y}_h) = MSE(X_h'(X'X)^{-1} X_h) \end{aligned}$$

- Prediction of New observation at X_h

$$s^2_{pred} = MSE(1 + X_h'(X'X)^{-1} X_h)$$

Analysis of Variance Results (ANOVA)

Total Sum of Squares

$$SSTO = \sum (Y_i - \bar{Y})^2 = \mathbf{Y}'\mathbf{Y} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}'[\mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J}]\mathbf{Y}$$

Proof:

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

$$= Y'Y - \frac{1}{n}(Y'1(Y'1)')$$

$$= Y'Y - \frac{1}{n}(Y'11'Y)$$

$$= Y'Y - \frac{1}{n}(Y'JY)$$

$$= Y'(I - \frac{1}{n}J)Y$$

Analysis of Variance Results (ANOVA) (cont.)

Error Sum of Squares

$$SSE = e'e = (\mathbf{Y}-\mathbf{Xb})'(\mathbf{Y}-\mathbf{Xb}) = \underbrace{\mathbf{Y}'\mathbf{Y}-\mathbf{b}'\mathbf{X}'\mathbf{Y}}_{①} = \underbrace{\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}}_{②}$$

Proof:

① $SSE = e'e = (Y-Xb)'(Y-Xb) = (Y'-b'X')(Y-Xb)$

$= Y'Y - Y'Xb - b'X'Y + b'X'Xb$, $\quad b = (X'X)^{-1}X'Y$

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

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

$= Y'Y - Y'Xb - b'X'Y + Y'Xb$

$= Y'Y - b'X'Y$

② $SSE = e'e = (Y-\hat{Y})'(Y-\hat{Y}) = ((I-H)Y)'(I-H)Y$

$= Y'(I-H)(I-H)Y$

$= Y'(I-H)Y \qquad$ Q.E.D.

Analysis of Variance Results (ANOVA) (cont.)

Regression Sum of Squares

$$SSR = \sum(\hat{Y}_i - \bar{Y})^2 = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}'[\mathbf{H} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J}]\mathbf{Y}$$

Proof:

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

$$= \sum_{i=1}^n \hat{Y}_i^2 - 2\bar{Y}\sum_{i=1}^n[\bar{Y} + b_1(X_i - \bar{X})] + n\bar{Y}^2, \quad \hat{Y}_i = b_0 + b_1X_i = \bar{Y} - b_1\bar{X} + b_1X_i = \bar{Y} + b_1(X_i - \bar{X})$$

$$= \sum_{i=1}^n \hat{Y}_i^2 - 2\bar{Y}[n\bar{Y} + \underbrace{b_1\sum(X_i - \bar{X})}_{=0}] + n\bar{Y}^2$$

$$= \sum_{i=1}^n \hat{Y}_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2 \boxed{= \sum_{i=1}^n \hat{Y}_i^2 - n\bar{Y}^2}$$

① $\sum \hat{Y}_i^2 = (Xb)^T(Xb) = b^TX^TXb = b'(X'X)\underbrace{((X'X)^{-1}X'Y)}_{b} = b'X'Y$

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

$SSR = ① - ② = b'X'Y - \frac{1}{n}Y'JY$

③ $\sum \hat{Y}_i^2 = \hat{Y}^T\hat{Y} = (HY)^T(HY) = Y'HY \Rightarrow SSR = ③ - ② = Y'(H - \frac{1}{n}J)Y$

Analysis of Variance Results (ANOVA) (Summary)

- Sums of Squares as Quadraatic Froms: $Y'AY$

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

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

$$SSR = \sum(\hat{Y}_i - \bar{Y})^2 = \mathbf{Y}'[\mathbf{H} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J}]\mathbf{Y}$$

Practice problems and upcoming topics

- Practice problems after today's lecture: CH5: 5.1, 5.3, 5.4, 5.8, 5.10, 5.12, 5.17, 5.21, 5.23, 5.29, 5.31
- Extra exercises
  - Show $I - H$ is symmetric and idempotent.
  - Show $H(I - H) = 0$
- Upcoming topics
  - CH6: Models with more than one predictors.
- Reading for upcoming topics: Chapter 6.