Lecture 9
SLR in Matrix Form

STAT 512
Spring 2011

Background Reading
KNNL: Chapter 5
Topic Overview

- Matrix Equations for SLR

- Don’t focus so much on the matrix arithmetic as on the form of the equations.

- Try to understand how the different pieces fit together.
SLR Model in Scalar Form

\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{where} \quad \begin{cases} \varepsilon_i & \overset{iid}{\sim} N(0, \sigma^2) \\ i = 1, 2, \ldots, n \end{cases} \]

Consider now writing an equation for each obsn:

\[
\begin{align*}
Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\
&\vdots \quad \vdots \\
Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n
\end{align*}
\]

9-3
SLR Model in Matrix Form

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

\[ \mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times 2} \mathbf{\beta}_{2 \times 1} + \varepsilon_{n \times 1} \]
Terminology

- $Y$ is the response vector
- $X$ is called the design matrix
- $\beta$ is the vector of parameters
- $\varepsilon$ is the error vector
Covariance Matrix for $\varepsilon$

$$\sigma^2 \{ \varepsilon \} = \text{Cov} \left[ \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \right] = \sigma^2 I_{n \times n} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$
Covariance Matrix for $Y$

$$\sigma^2 \{ Y \} = \text{Cov} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sigma^2 I_{n \times n}$$
Assumptions in Matrix Form

- $\varepsilon \sim N(0, \sigma^2 I)$
- 0 is the $n \times 1$ zero vector; $I$ is the $n \times n$ identity matrix.
- Ones in the diagonal elements specify that the variance of each $\varepsilon_i$ is $\sigma^2$.
- Zeros in the off-diagonal elements specify that the covariance between $\varepsilon_i$ and $\varepsilon_j$ is zero for $i \neq j$.
- Implies zero correlation.
Least Squares Estimation

Errors are $\varepsilon = Y - X\beta$. Want to minimize sum of squared errors:

$$\sum \varepsilon_i^2 = \left[ \varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n \right] \left[ \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{array} \right] = \varepsilon'\varepsilon$$

That is, we want to minimize:

$$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$
Least Squares (2)

We take the derivative with respect to the vector $\beta$

- This is like a quadratic function: think “$(Y - X\beta)^2$”.

- Using the chain rule, this derivative works out to

$$\frac{d}{d\beta} \left( (Y - X\beta)' (Y - X\beta) \right) = -2X' (Y - X\beta)$$
Least Squares (3)

We set this equal to 0 (a vector of zeros) and solve for $\beta$, resulting in the normal equations:

$$X'Y = (X'X)\beta$$

Solving this equation for $\beta$ gives the least squares solution for $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X'X)^{-1}(X'Y).$
Least Squares Solution

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

(REMEMBER THIS!!!)
Reality break

This is just to convince you that we have done nothing new or magical – all we are doing is writing the same old formulas for $b_0$ and $b_1$ in matrix format. Do NOT worry if you cannot reproduce the following matrix algebra, but you SHOULD try to follow it so that you believe me that this is really not a new formula.
Recall from previous topics we had:

\[
b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \triangleq \frac{SS_{XY}}{SS_X}
\]

\[
b_0 = \bar{Y} - b_1 \bar{X}
\]

For the matrix format, we begin by noting:

\[
X'X = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
X_1 & X_2 & \cdots & X_n \\
1 & X_2 & \cdots & X_n \\
\vdots & \vdots & \ddots & \vdots \\
1 & X_n & \cdots & 1
\end{bmatrix}\begin{bmatrix}
1 \\
1 \\
1 \\
\vdots \\
1
\end{bmatrix} = \begin{bmatrix}
n & \sum X_i \\
\sum X_i & \sum X_i^2
\end{bmatrix}
\]
\[
(X'X)^{-1} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix}
\sum X_i^2 & -\sum X_i \\
-\sum X_i & n
\end{bmatrix}
\]

\[
= \frac{1}{nSS_X} \begin{bmatrix}
\sum X_i^2 & -\sum X_i \\
-\sum X_i & n
\end{bmatrix}
\]

\[
X'Y = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
X_1 & X_2 & \cdots & X_n
\end{bmatrix} \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{bmatrix} = \begin{bmatrix}
\sum Y_i \\
\sum X_iY_i
\end{bmatrix}
\]

Plugging these into the equation for \( b \) yields:
\[
\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}
\]

\[
= \frac{1}{nSS_X} \begin{bmatrix}
\sum X_i^2 & -\sum X_i \\
-\sum X_i & n
\end{bmatrix} \begin{bmatrix}
\sum Y_i \\
\sum X_i Y_i
\end{bmatrix}
\]

\[
= \frac{1}{nSS_X} \begin{bmatrix}
(\sum X_i^2)(\sum Y_i) - (\sum X_i)(\sum X_i Y_i) \\
-(\sum X_i)(\sum Y_i) + n \sum X_i Y_i
\end{bmatrix}
\]

\[
= \frac{1}{SS_X} \begin{bmatrix}
\bar{Y} (\sum X_i^2) - \bar{X} \sum X_i Y_i \\
\sum X_i Y_i - n \bar{X} \bar{Y}
\end{bmatrix}
\]

\[
= \frac{1}{SS_X} \begin{bmatrix}
\bar{Y} (\sum X_i^2) - \bar{Y} (n \bar{X}^2) + \bar{X} (n \bar{X} \bar{Y}) - \bar{X} \sum X_i Y_i \\
SS_{XY}
\end{bmatrix}
\]

\[
= \frac{1}{SS_X} \begin{bmatrix}
\bar{Y} SS_X - SS_{XY} \bar{X} \\
SS_{XY}
\end{bmatrix} = \begin{bmatrix}
\bar{Y} - \frac{SS_{XY}}{SS_X} \bar{X} \\
\frac{SS_{XY}}{SS_X}
\end{bmatrix} = \begin{bmatrix}
b_0 \\
b_1
\end{bmatrix}
\]
SLR in Matrices

- All we have done is to write the same old formulas for $b_0$ and $b_1$ in a fancy new format.

- Why have we bothered to do this? The cool part is that the same approach works for multiple regression. All we do is make $X$ and $b$ into bigger matrices, and use exactly the same formula.
Fitted Values

- The fitted (or predicted) values are:

\[
\hat{Y} = \begin{bmatrix}
\hat{Y}_1 \\
\hat{Y}_2 \\
\vdots \\
\hat{Y}_n
\end{bmatrix} = \begin{bmatrix}
b_0 + b_1 X_1 \\
b_0 + b_1 X_2 \\
\vdots \\
b_0 + b_1 X_n
\end{bmatrix} = \begin{bmatrix}
1 & X_1 \\
1 & X_2 \\
\vdots & \vdots \\
1 & X_n
\end{bmatrix} \begin{bmatrix}
b_0 \\
b_1
\end{bmatrix} = Xb
\]
Hat Matrix

\[ \hat{Y} = Xb \]

\[ \hat{Y} = X(X'X)^{-1}X'Y \]

\[ \hat{Y} = HY \quad \text{where} \quad H = X(X'X)^{-1}X' \]

- Involved in standard errors
- Plays large role in diagnostics and the identification of influential observations
Hat Matrix Properties

- Hat matrix is symmetric
- Hat matrix is *idempotent* since

\[ HH = H \]
Residuals

• Formula: \( e = Y - Xb = (I - H)Y \)

• Only \( Y \) is random. \( Var(Y) = \sigma^2 I \)

• Variance Covariance Matrix for \( e \) is the \( n \times n \) matrix:

\[
Var \{ e \} = (I - H)^2 \sigma^2 I = \sigma^2 (I - H)
\]

• Estimated covariance matrix for \( e \) is (also an \( n \times n \) matrix):

\[
s^2 \{ e \} = MSE (I - H)
\]
Other Variances / Covariances

- The vector $b$ is a linear combination of the elements of $Y$.
- Since $Y$ is normal, $b$ must also be normal.
- Even if normality is violated, these estimates are robust and will be approximately normal in general.
- To determine the var-cov matrix we need to use the following theorem:
Useful Multivariate Theorem

Suppose \( U \sim N(\mu, \Sigma) \), a multivariate normal vector, and \( V = c + DU \), a linear transformation of \( U \) where \( c \) is a vector and \( D \) is a matrix.

Then \( V \sim N(c + D\mu, D\Sigma D') \).
Covariance Matrix for \( b \)

- Recall: \( b = (X'X)^{-1} X'Y = [(X'X)^{-1} X'] Y \)
  
  and \( Y \sim \mathcal{N}(X\beta, \sigma^2 I) \).

- Now apply theorem to \( b \) using

\[
U = Y, \quad \mu = X\beta, \quad \Sigma = \sigma^2 I \\
V = b, \quad c = 0 \text{ and } D = (X'X)^{-1} X'
\]
Covariance Matrix for $b$ (2)

The theorem tells us the vector $b$ is normally distributed with mean

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

and covariance matrix

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

Note: Use basic results from section 5.7 and the fact that both $X'X$ and its inverse are symmetric to show this equality.
Other Variances in Matrix Form

\[
s^2 \{ \hat{Y}_h \} = MSE \left( X_h' (X'X)^{-1} X_h \right)
\]

\[
s^2 \{ \hat{Y}_{h,new} \} = MSE \left( 1 + X_h' (X'X)^{-1} X_h \right)
\]

Note: \( X_h' \) is the row vector \([1 \ \ X_h] \).
Sums of Squares

- Sums of squares can also be written in terms of matrices.

- Need a special matrix $J$ which is square and every entry is 1.

- Results shown in Section 5.12.
Sum of Squares Formulas

\[ SSR = b'X'Y - \left( \frac{1}{n} \right) Y'JY \]

\[ SSE = e'e = Y'Y - b'X'Y \]

\[ SSTO = Y'Y - \left( \frac{1}{n} \right) Y'JY \]
Sums of Squares (2)

- Can also be written as *quadratic forms*

\[
SSR = Y' \left[ H - \left( \frac{1}{n} \right) J \right] Y
\]

\[
SSE = Y' [ I - H ] Y
\]

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

- As usual we will continue to use software to do computations, so you need not worry about having to multiply big matrices.

- “Thinking” in matrix form will come in handy when we begin talking about multiple regression, because all one needs to do is extend the dimensions of the matrices.
Upcoming in Lecture 10...

- Introduction to Multiple Regression
- Background Reading:
  - KNNL 6.1-6.3