Matrix Approach to Linear Regression

Dr. Frank Wood
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

\[ \mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} \mathbf{E}\{Y_1\} \\ \mathbf{E}\{Y_2\} \\ \mathbf{E}\{Y_3\} \end{bmatrix} \]
Expectation of a Random Matrix

- The expectation of a random matrix is defined similarly

\[
E\{Y\} = [E\{Y_{ij}\}]_{n \times p} \quad i = 1, \ldots, n; j = 1, \ldots, p
\]
Covariance Matrix of a Random Vector

- The collection of variances and covariances of and between the elements of a random vector can be collection into a matrix called the covariance matrix

\[
\sigma^2\{Y\} = \begin{bmatrix}
\sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\
\sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_3\} \\
\sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\}
\end{bmatrix}
\]

remember

\[
\sigma\{Y_2, Y_1\} = \sigma\{Y_1, Y_2\}
\]

so the covariance matrix is symmetric
Derivation of Covariance Matrix

• In vector terms the covariance matrix is defined by

\[ \sigma^2 \{ \mathbf{Y} \} = \mathbf{E} \left\{ (\mathbf{Y} - \mathbf{E} \{ \mathbf{Y} \}) (\mathbf{Y} - \mathbf{E} \{ \mathbf{Y} \})' \right\} \]

because

\[ \sigma^2 \{ \mathbf{Y} \} = \mathbf{E} \left\{ \begin{bmatrix} Y_1 - \mathbf{E} \{ Y_1 \} \\ Y_2 - \mathbf{E} \{ Y_2 \} \\ Y_3 - \mathbf{E} \{ Y_3 \} \end{bmatrix} \begin{bmatrix} Y_1 - \mathbf{E} \{ Y_1 \} & Y_2 - \mathbf{E} \{ Y_2 \} & Y_3 - \mathbf{E} \{ Y_3 \} \end{bmatrix} \right\} \]

verify first entry
Regression Example

• Take a regression example with \( n=3 \) with constant error terms \( \sigma^2\{\epsilon_i\} = \sigma^2 \) and are uncorrelated so that \( \sigma^2\{\epsilon_i, \epsilon_j\} = 0 \) for all \( i \neq j \)

• The covariance matrix for the random vector \( \epsilon \) is

\[
\sigma^2\{\epsilon\} = \begin{bmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & 0 \\
0 & 0 & \sigma^2 \\
\end{bmatrix}_{3 \times 3}
\]

which can be written as

\[
\sigma^2\{\epsilon\} = \sigma^2 I_{3 \times 3}
\]
Basic Results

• If $A$ is a constant matrix and $Y$ is a random matrix then

$$W = AY$$

is a random matrix

$$E\{A\} = A$$

$$E\{W\} = E\{AY\} = AE\{Y\}$$

$$\sigma^2\{W\} = \sigma^2\{AY\} = A\sigma^2\{Y\}A'$$
Multivariate Normal Density

• Let $Y$ be a vector of $p$ observations

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}_{p \times 1}$$

• Let $\mu$ be a vector of $p$ means for each of the $p$ observations

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}_{p \times 1}$$
Multivariate Normal Density

- Let $\Sigma$ be the covariance matrix of $Y$

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\
\sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2
\end{bmatrix}
\]

- Then the multivariate normal density is given by

\[
f(Y) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (Y - \mu)' \Sigma^{-1} (Y - \mu) \right]
\]
Example 2d Multivariate Normal Distribution

Run multivariate_normal_plots.m
Matrix Simple Linear Regression

• Nothing new – only matrix formalism for previous results
• Remember the normal error regression model

\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \ldots, n \]

• This implies

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

• If we identify the following matrices

\[
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n \times 2}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2 \times 1}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1}
\]

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

\[
Y = X \beta + \epsilon
\]

Frank Wood, fwood@stat.columbia.edu  Linear Regression Models  Lecture 11, Slide 12
Regression Matrices

• Of course, in the normal regression model the expected value of each of the $\varepsilon_i$’s is zero, we can write

$$E\{Y\} = X\beta$$

• This is because

$$E\{\varepsilon\} = \begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \vdots \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
Error Covariance

• Because the error terms are independent and have constant variance $\sigma^2$

$$\sigma^2 \{ \varepsilon \} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}_{n \times n}$$

$$\sigma^2 \{ \varepsilon \} = \sigma^2 \mathbf{I}_{n \times n}$$
Matrix Normal Regression Model

- In matrix terms the normal regression model can be written as

\[ Y = X\beta + \epsilon \]

where

\[ \mathbb{E}\{\epsilon\} = 0 \]

and

\[ \sigma^2\{\epsilon\} = \sigma^2 I \]
Least Squares Estimation

• Starting from the normal equations you have derived

\[ nb_0 + b_1 \sum X_i = \sum Y_i \]
\[ b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i \]

we can see that these equations are equivalent to the following matrix operations

\[
\begin{pmatrix}
X'X & b
\end{pmatrix}_{2 \times 2} \begin{pmatrix}
b_0 \\
b_1
\end{pmatrix}_{2 \times 1} =
\begin{pmatrix}
X'Y
\end{pmatrix}_{2 \times 1}
\]

with

\[
\begin{pmatrix}
b_0 \\
b_1
\end{pmatrix}_{2 \times 1}
\]

demonstrate this on board
Estimation

• We can solve this equation

\[ X'X \begin{bmatrix} b \\ 1 \end{bmatrix} = X'Y \]

(if the inverse of \( X'X \) exists) by the following

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

and since

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

we have

\[ b = (X'X)^{-1}X'Y \]
Least Squares Solution

• The matrix normal equations can be derived directly from the minimization of

\[ Q = (Y - X\beta)'(Y - X\beta) \]

w.r.t. to \( \beta \)

Do this on board.
Fitted Values and Residuals

• Let the vector of the fitted values be

\[
\hat{\mathbf{Y}} = \begin{bmatrix}
\hat{Y}_1 \\
\hat{Y}_2 \\
\vdots \\
\hat{Y}_n
\end{bmatrix}
\]

in matrix notation we then have

\[
\hat{\mathbf{Y}} = \mathbf{X} \mathbf{b}
\]
Hat Matrix – Puts hat on Y

- We can also directly express the fitted values in terms of only the X and Y matrices

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

and we can further define H, the “hat matrix”

\[ \hat{Y} \in n \times 1 \]
\[ H \in n \times n \]
\[ Y \in n \times 1 \]

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

- The hat matrix plans an important role in diagnostics for regression analysis.
Hat Matrix Properties

- The hat matrix is symmetric
- The hat matrix is idempotent, i.e.

\[ HH = H \]

demonstrate on board
Residuals

• The residuals, like the fitted values of $\hat{Y}_i$ can be expressed as linear combinations of the response variable observations $Y_i$

\[
e = Y - \hat{Y} = Y - HY = (I - H)Y
\]
Covariance of Residuals

- Starting with \( e = (I - H)Y \),

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

\[
\sigma^2\{Y\} = \sigma^2\{e\} = \sigma^2 I
\]

which means that

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

\[
= \sigma^2 (I - H)(I - H)
\]

and since \( I - H \) is idempotent (check) we have

\[
\sigma^2\{e\} = \sigma^2 (I - H)_{n \times n}
\]

we can plug in MSE for \( \sigma^2 \) as an estimate
ANOVA

- We can express the ANOVA results in matrix form as well, starting with

\[ SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \]

where

\[ Y'Y = \sum Y_i^2 \]

\[ \frac{(\sum Y_i)^2}{n} = \left( \frac{1}{n} \right) Y'JY \]

leaving

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

J is matrix of all ones, do 3x3 example
SSE

- Remember

\[ SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \]

- We have

\[ SSE = e'e = (Y - Xb)'(Y - Xb) = Y'Y - 2b'X'Y + b'X'Xb \]

derive this on board

\[ SSE = Y'Y - 2b'X'Y + b'X'X(X'X)^{-1}X'Y \]
\[ = Y'Y - 2b'X'Y + b'IX'Y \]

- Simplified

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

- It can be shown that
  - for instance, remember $SSR = \text{SSTO} - \text{SSE}$

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

\[
\text{SSTO} = Y'Y - \left(\frac{1}{n}\right)Y'JY
\quad \text{SSE} = Y'Y - b'X'Y
\]

write these on board
Tests and Inference

• The ANOVA tests and inferences we can perform are the same as before
• Only the algebraic method of getting the quantities changes
• Matrix notation is a writing short-cut, not a computational shortcut
Quadratic Forms

• The ANOVA sums of squares can be shown to be quadratic forms. An example of a quadratic form is given by

\[ 5Y_1^2 + 6Y_1Y_2 + 4Y_2^2 \]

• Note that this can be expressed in matrix notation as (where A is a symmetric matrix)

\[ \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = Y'AY \]
Quadratic Forms

• In general, a quadratic form is defined by

\[ Y'AY = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} Y_i Y_j \quad \text{where } a_{ij} = a_{ji} \]

A is the matrix of the quadratic form.

• The ANOVA sums SSTO, SSE, and SSR are all quadratic forms.
ANOVA quadratic forms

• Consider the following reexpression of $b'X'$

$$b'X' = (Xb)' = \hat{Y}' \quad b'X' = (HY)' \quad b'X' = Y'H$$

• With this it is easy to see that

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

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

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

• We can derive the sampling variance of the $\beta$ vector estimator by remembering that

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

where $A$ is a constant matrix

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

which yields

$$\sigma^2\{b\} = A\sigma^2\{Y\}A'$$
Variance of $b$

- Since $(X'X)^{-1}$ is symmetric we can write

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

and thus

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

- Of course this assumes that we know $\sigma^2$. If we don’t, we, as usual, replace it with the MSE.

\[
\sigma^2\{b\}_{2\times2} = \begin{bmatrix}
\frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum(X_i - \bar{X})^2} & -\bar{X}\sigma^2 \\
-\bar{X}\sigma^2 & \frac{\sigma^2}{\sum(X_i - \bar{X})^2}
\end{bmatrix}
\]

\[
s^2\{b\} = MSE(X'X)^{-1} = \begin{bmatrix}
\frac{MSE}{n} + \frac{\bar{X}^2MSE}{\sum(X_i - \bar{X})^2} & -\bar{X}MSE \\
-\bar{X}MSE & \frac{MSE}{\sum(X_i - \bar{X})^2}
\end{bmatrix}
\]
Mean Response

• To estimate the mean response we can create the following matrix

\[
X_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \quad \text{or} \quad X'_h = \begin{bmatrix} 1 & X_h \end{bmatrix}
\]

• The fit (or prediction) is then

\[
\hat{Y}_h = X'_h b
\]

since

\[
X'_h b = \begin{bmatrix} 1 & X_h \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h] = [\hat{Y}_h] = \hat{Y}_h
\]
Variance of Mean Response

- Is given by

\[ \sigma^2 \{ \hat{Y}_h \} = \sigma^2 X_h' (X'X)^{-1} X_h \]

and is arrived at in the same way as for the variance of \( \beta \)

- Similarly the estimated variance in matrix notation is given by

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

• Expectation and variance of random vector and matrices
• Simple linear regression in matrix form
• Next: multiple regression