2.8 Matrix approach to simple linear regression

In this section we will briefly discuss a matrix approach to fitting simple linear regression models. A random sample of size \( n \) gives \( n \) equations. For the full SLRM we have

\[
Y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\
Y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2 \\
\vdots \\
Y_n = \beta_0 + \beta_1 x_n + \varepsilon_n
\]

We can write this in matrix formulation as

\[
Y = X\beta + \varepsilon, \quad (2.22)
\]

where \( Y \) is an \((n \times 1)\) vector of response variables (random sample), \( X \) is an \((n \times 2)\) matrix called the design matrix, \( \beta \) is a \((2 \times 1)\) vector of unknown parameters and \( \varepsilon \) is an \((n \times 1)\) vector of random errors. That is,

\[
Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.
\]

The assumptions about the random errors let us write

\[
\varepsilon \sim \mathcal{N}_n \left( 0, \sigma^2 I \right),
\]

that is vector \( \varepsilon \) has \( n \)-dimensional normal distribution with

\[
E(\varepsilon) = E\left( \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right) = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0
\]

and the variance-covariance matrix

\[
\text{Var}(\varepsilon) = \begin{pmatrix} \text{var}(\varepsilon_1) & \text{cov}(\varepsilon_1, \varepsilon_2) & \ldots & \text{cov}(\varepsilon_1, \varepsilon_n) \\ \text{cov}(\varepsilon_2, \varepsilon_1) & \text{var}(\varepsilon_2) & \ldots & \text{cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\varepsilon_n, \varepsilon_1) & \text{cov}(\varepsilon_n, \varepsilon_2) & \ldots & \text{var}(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix} = \sigma^2 I
\]
This formulation is usually called the Linear Model (in $\beta$). All the models we have considered so far can be written in this general form. The dimensions of matrix $X$ and of vector $\beta$ depend on the number $p$ of parameters in the model and, respectively, they are $n \times p$ and $p \times 1$. In the full SLRM we have $p = 2$.

The null model ($p = 1$)

$$Y_i = \beta_0 + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n$$

is equivalent to

$$Y = 1\beta_0 + \varepsilon$$

where $1$ is an $(n \times 1)$ vector of $1$’s.

The no-intercept model ($p = 1$)

$$Y_i = \beta_1 x_i + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n$$

can be written as in matrix notation with

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \end{pmatrix}.$$ 

Quadratic regression ($p = 3$)

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n$$

can be written in matrix notation with

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

The normal equations obtained in the least squares method are given by

$$X^T Y = X^T \hat{X} \hat{\beta}.$$
It follows that so long as \( X^T X \) is invertible, i.e., its determinant is non-zero, the unique solution to the normal equations is given by

\[
\hat{\beta} = (X^T X)^{-1} X^T Y.
\]

This is a common formula for all linear models where \( X^T X \) is invertible. For the full simple linear regression model we have

\[
X^T Y = \begin{pmatrix}
1 & 1 & \cdots & 1 \\
x_1 & x_2 & \cdots & x_n
\end{pmatrix}
\begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{pmatrix}
= \begin{pmatrix}
\sum Y_i \\
\sum x_i Y_i
\end{pmatrix}
= \begin{pmatrix}
n \bar{Y} \\
\sum x_i Y_i
\end{pmatrix}
\]

and

\[
X^T X = \begin{pmatrix}
n \\
\sum x_i \\
\sum x_i^2
\end{pmatrix}
= \begin{pmatrix}
n \bar{x} \\
\sum x_i \\
\sum x_i^2
\end{pmatrix}.
\]

The determinant of \( X^T X \) is given by

\[
|X^T X| = n \sum x_i^2 - (n \bar{x})^2 = n \left( \sum x_i^2 - n \bar{x}^2 \right) = n S_{xx}.
\]

Hence, the inverse of \( X^T X \) is

\[
(X^T X)^{-1} = \frac{1}{n S_{xx}} \begin{pmatrix}
\sum x_i^2 - n \bar{x}^2 \\
-n \bar{x}^2 \\
n \bar{x}
\end{pmatrix}
= \frac{1}{S_{xx}} \begin{pmatrix}
\frac{1}{n} \sum x_i^2 - \bar{x}^2 \\
-\bar{x}^2 \\
\bar{x}
\end{pmatrix}.
\]

So the solution to the normal equations is given by

\[
\hat{\beta} = (X^T X)^{-1} X^T y
\]

\[
= \frac{1}{S_{xx}} \begin{pmatrix}
\frac{1}{n} \sum x_i^2 - \bar{x}^2 \\
-\bar{x}^2 \\
\bar{x}
\end{pmatrix}
\begin{pmatrix}
\sum x_i Y_i \\
\sum x_i^2 Y_i \\
\sum x_i Y_i^2
\end{pmatrix}
\]

\[
= \frac{1}{S_{xx}} \begin{pmatrix}
\bar{Y} \sum x_i^2 - \bar{x} \sum x_i Y_i \\
\sum x_i Y_i - n \bar{x} \bar{Y} \\
\sum x_i Y_i - n \bar{x} \sum x_i Y_i
\end{pmatrix}
\]

\[
= \frac{1}{S_{xx}} \begin{pmatrix}
\bar{Y} \sum x_i^2 - \bar{x}^2 \bar{Y} + n \bar{x}^2 \bar{Y} - \bar{x} \sum x_i Y_i \\
\sum x_i^2 Y_i - n \bar{x} \bar{Y} \\
\sum x_i Y_i - n \bar{x} \sum x_i Y_i
\end{pmatrix}
\]

\[
= \frac{1}{S_{xx}} \begin{pmatrix}
\bar{Y} S_{xx} - \bar{x} S_{xy} \\
\sum x_i Y_i - \bar{x} \sum x_i Y_i \\
\sum x_i Y_i - n \bar{x} \sum x_i Y_i
\end{pmatrix}
\]

\[
= \left( \begin{array}{c}
\bar{Y} - \hat{\beta}_1 \bar{x} \\
\hat{\beta}_1 \\
\end{array} \right)
\]
which is the same result as we obtained before.

Note:
Let \( A \) and \( B \) be a vector and a matrix of real constants and let \( Z \) be a vector of random variables, all of appropriate dimensions so that the addition and multiplication are possible. Then

\[
\begin{align*}
E(A + BZ) &= A + BE(Z) \\
\text{Var}(A + BZ) &= \text{Var}(BZ) = B \text{Var}(Z)B^T.
\end{align*}
\]

In particular,

\[
\begin{align*}
E(Y) &= E(X\beta + \varepsilon) = X\beta \\
\text{Var}(Y) &= \text{Var}(X\beta + \varepsilon) = \text{Var}(\varepsilon) = \sigma^2 I.
\end{align*}
\]

These equalities let us prove the following theorem.

**Theorem 2.7.** The least squares estimator \( \hat{\beta} \) of \( \beta \) is unbiased and its variance-covariance matrix is

\[ \text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}. \]

**Proof.** First we will show that \( \hat{\beta} \) is unbiased. Here we have

\[
\begin{align*}
E(\hat{\beta}) &= E\{(X^T X)^{-1} X^T Y\} = (X^T X)^{-1} X^T E(Y) \\
&= (X^T X)^{-1} X^T X\beta = I\beta = \beta.
\end{align*}
\]

Now, we will show the result for the variance-covariance matrix.

\[
\begin{align*}
\text{Var}(\hat{\beta}) &= \text{Var}\{(X^T X)^{-1} X^T Y\} \\
&= (X^T X)^{-1} X^T \text{Var}(Y) X(X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} X^T I X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.
\end{align*}
\]

We denote the vector of residuals as

\[ e = Y - \hat{Y}, \]

where \( \hat{Y} = \hat{E}(Y) = X\hat{\beta} \) is the vector of fitted responses \( \hat{\mu}_i \). It can be shown that the following theorem holds.
Theorem 2.8. The \( n \times 1 \) vector of residuals \( e \) has mean
\[
E(e) = 0
\]
and variance-covariance matrix
\[
\text{Var}(e) = \sigma^2 (I - X(X^T X)^{-1}X^T).
\]

Hence, variance of the residuals \( e_i \) is
\[
\text{var}[e_i] = \sigma^2 (1 - h_{ii}),
\]
where the leverage \( h_{ii} \) is the \( i \)th diagonal element of the Hat Matrix
\[ H = X(X^T X)^{-1}X^T, \]
i.e.,
\[
h_{ii} = x_i^T(X^T X)^{-1}x_i,
\]
where \( x_i^T = (1, x_i) \) is the \( i \)th row of matrix \( X \).

The \( i \)th mean response can be written as
\[
E(Y_i) = \mu_i = x_i^T \beta = (1, x_i) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0 + \beta_1 x_i
\]
and its estimator as
\[
\hat{\mu}_i = x_i^T \hat{\beta}.
\]
Then, the variance of the estimator is
\[
\text{var}(\hat{\mu}_i) = \text{var}(x_i^T \hat{\beta}) = \sigma^2 x_i^T (X^T X)^{-1}x_i = \sigma^2 h_{ii}
\]
and the estimator of this variance is
\[
\hat{\text{var}}(\hat{\mu}_i) = S^2 h_{ii},
\]
where \( S^2 \) is a suitable unbiased estimator of \( \sigma^2 \).

We can easily obtain other results we have seen for the SLRM written in non-matrix notation, now using the matrix notation, both for the full model and for a reduced SLM (no intercept or zero slope).

We have seen on page 50 that
\[
(X^T X)^{-1} = \frac{1}{nS_{xx}^2} \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}.
\]
Now, by Theorem 2.7, \( \text{Var}[\hat{\beta}] = \sigma^2 (X^T X)^{-1} \). Thus

\[
\text{var}[\hat{\beta}_0] = \frac{\sigma^2}{nS_{xx}} \sum x_i^2
\]

which, by writing \( \sum x^2 = \sum x^2 - n\bar{x}^2 + n\bar{x}^2 \), can be written as \( \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right\} \).

Also,

\[
\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \left( -\frac{n\bar{x}}{nS_{xx}} \right)
\]

and

\[
\text{var}[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}}.
\]

The quantity \( h_{ii} \) is given by

\[
h_{ii} = x_i^T (X^T X)^{-1} x_i
\]

\[
= (1 x_i) \frac{1}{nS_{xx}} \left( \sum x_j^2 - n\bar{x} \right) \left( 1 \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} \right)
\]

We shall leave it as an exercise to show that this simplifies to

\[
h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}.
\]

### 2.8.1 Some specific examples

1. The Null model

As we have seen, this can be written as

\[
Y = X \beta_0 + \epsilon
\]

where \( X = 1 \) is an \( (n \times 1) \) vector of 1’s. So \( X^T X = n, \ X^T Y = \sum Y_i \), which gives

\[
\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{1}{n} \sum Y_i = \bar{Y} = \hat{\beta}_0,
\]

\[
\text{var}[\hat{\beta}] = (X^T X)^{-1} \sigma^2 = \frac{\sigma^2}{n}.
\]
2. No-intercept model

We saw that this example fits the General Linear Model with

\[
X = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{bmatrix}, \quad \beta = \beta_1
\]

So \( X^T X = \sum x_i^2 \) and \( X^T Y = \sum x_i Y_i \), and we can calculate

\[
\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{\sum x_i Y_i}{\sum x_i^2} = \hat{\beta}_1,
\]

\[
\text{Var}[\hat{\beta}] = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{\sum x_i^2}.
\]