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

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots & \qquad \vdots \\ Y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n \end{aligned}$$

We can write this in matrix formulation as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.22}$$

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

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

The assumptions about the random errors let us write

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

that is vector $\boldsymbol{\varepsilon}$ has n-dimensional normal distribution with

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathrm{E}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \mathrm{E}(\varepsilon_1) \\ \mathrm{E}(\varepsilon_2) \\ \vdots \\ \mathrm{E}(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{0}$$

and the variance-covariance matrix

$$\begin{aligned} \mathrm{Var}(\boldsymbol{\varepsilon}) &= \begin{pmatrix} \mathrm{var}(\varepsilon_1) & \mathrm{cov}(\varepsilon_1, \varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_1, \varepsilon_n) \\ \mathrm{cov}(\varepsilon_2, \varepsilon_1) & \mathrm{var}(\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\varepsilon_n, \varepsilon_1) & \mathrm{cov}(\varepsilon_n, \varepsilon_2) & \dots & \mathrm{var}(\varepsilon_n) \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I} \end{aligned}$$

This formulation is usually called the *Linear Model* (in $\boldsymbol{\beta}$). All the models we have considered so far can be written in this general form. The dimensions of matrix $\boldsymbol{X}$ and of vector $\boldsymbol{\beta}$ 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 $(p = 1)$

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

is equivalent to

$$\boldsymbol{Y} = \mathbf{1}\beta_0 + \boldsymbol{\varepsilon}$$

where $\mathbf{1}$ is an $(n \times 1)$ vector of 1's.

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

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

can be written as in matrix notation with

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

Quadratic regression $(p = 3)$

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

can be written in matrix notation with

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

□

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

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

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

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}.$$

This is a common formula for all linear models where $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is invertible. For the full simple linear regression model we have

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

and

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}.$$

The determinant of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is given by

$$|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}| = n\sum x_i^2 - (n\bar{x})^2 = n\left(\sum x_i^2 - n\bar{x}^2\right) = nS_{xx}.$$

Hence, the inverse of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \frac{1}{nS_{xx}} \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} = \frac{1}{S_{xx}} \begin{pmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

So the solution to the normal equations is given by

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}y \\
&= \frac{1}{S_{xx}} \begin{pmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} n\bar{Y} \\ \sum x_i Y_i \end{pmatrix} \\
&= \frac{1}{S_{xx}} \begin{pmatrix} \bar{Y}\sum x_i^2 - \bar{x}\sum x_i Y_i \\ \sum x_i Y_i - n\bar{x}\bar{Y} \end{pmatrix} \\
&= \frac{1}{S_{xx}} \begin{pmatrix} \bar{Y}\sum x_i^2 - n\bar{x}^2\bar{Y} + n\bar{x}^2\bar{Y} - \bar{x}\sum x_i Y_i \\ S_{xY} \end{pmatrix} \\
&= \frac{1}{S_{xx}} \begin{pmatrix} \bar{Y}(\sum x_i^2 - n\bar{x}^2) - \bar{x}(\sum x_i Y_i - n\bar{x}\bar{Y}) \\ S_{xY} \end{pmatrix} \\
&= \frac{1}{S_{xx}} \begin{pmatrix} \bar{Y}S_{xx} - \bar{x}S_{xY} \\ S_{xY} \end{pmatrix} \\
&= \begin{pmatrix} \bar{Y} - \widehat{\beta}_1\bar{x} \\ \widehat{\beta}_1 \end{pmatrix}
\end{aligned}$$

which is the same result as we obtained before. □

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

$$\begin{aligned} \mathrm{E}(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{Z}) &= \boldsymbol{A} + \boldsymbol{B}\,\mathrm{E}(\boldsymbol{Z}) \\ \mathrm{Var}(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{Z}) &= \mathrm{Var}(\boldsymbol{B}\boldsymbol{Z}) = \boldsymbol{B}\,\mathrm{Var}(\boldsymbol{Z})\boldsymbol{B}^{\mathrm{T}}. \end{aligned}$$

In particular,

$$\begin{aligned} \mathrm{E}(\boldsymbol{Y}) &= \mathrm{E}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{\beta} \\ \mathrm{Var}(\boldsymbol{Y}) &= \mathrm{Var}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}. \end{aligned}$$

These equalities let us prove the following theorem.

Theorem 2.7. *The least squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is unbiased and its variance-covariance matrix is*

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}.$$

Proof. First we will show that $\widehat{\boldsymbol{\beta}}$ is unbiased. Here we have

$$\begin{aligned} \mathrm{E}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}\} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\,\mathrm{E}(\boldsymbol{Y}) \\ &= (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{I}\boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned}$$

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

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{Var}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}\} \\ &= (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\,\mathrm{Var}(\boldsymbol{Y})\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}. \end{aligned}$$

□

We denote the vector of residuals as

$$\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}},$$

where $\widehat{\boldsymbol{Y}} = \widehat{\mathrm{E}(\boldsymbol{Y})} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ is the vector of fitted responses $\widehat{\mu}_i$. It can be shown that the following theorem holds.

Theorem 2.8. *The $n \times 1$ vector of residuals $\boldsymbol{e}$ has mean*

$$E(\boldsymbol{e}) = \boldsymbol{0}$$

and variance-covariance matrix

$$\operatorname{Var}(\boldsymbol{e}) = \sigma^2\big(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\big).$$

□

Hence, variance of the residuals e_i is

$$\operatorname{var}[e_i] = \sigma^2(1 - h_{ii}),$$

where the leverage h_{ii} is the ith diagonal element of the *Hat Matrix* $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$, i.e.,

$$h_{ii} = \boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}_i,$$

where $\boldsymbol{x}_i^{\mathrm{T}} = (1, x_i)$ is the ith row of matrix $\boldsymbol{X}$.

The ith mean response can be written as

$$\mathrm{E}(Y_i) = \mu_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} = (1, x_i)\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0 + \beta_1 x_i$$

and its estimator as

$$\widehat{\mu}_i = \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}.$$

Then, the variance of the estimator is

$$\operatorname{var}(\widehat{\mu}_i) = \operatorname{var}(\boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}_i = \sigma^2 h_{ii}$$

and the estimator of this variance is

$$\widehat{\operatorname{var}(\widehat{\mu}_i)} = S^2 h_{ii},$$

where S^2 is a suitable unbiased estimator of σ^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

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \frac{1}{nS_{xx}}\begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}.$$

Now, by Theorem 2.7, $\text{Var}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}^{\text{T}}\boldsymbol{X})^{-1}$. Thus

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

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,

$$\begin{aligned} \text{cov}(\widehat{\beta}_0, \widehat{\beta}_1) &= \sigma^2 \left(\frac{-n\bar{x}}{nS_{xx}} \right) \\ &= \frac{-\sigma^2 \bar{x}}{S_{xx}}, \end{aligned}$$

and

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

The quantity h_{ii} is given by

$$\begin{aligned} h_{ii} &= \boldsymbol{x}_i^{\text{T}}(\boldsymbol{X}^{\text{T}}\boldsymbol{X})^{-1}\boldsymbol{x}_i \\ &= (1\, x_i) \frac{1}{nS_{xx}} \begin{pmatrix} \sum x_j^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix}. \end{aligned}$$

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

   $$\boldsymbol{Y} = \boldsymbol{X}\beta_0 + \boldsymbol{\varepsilon}$$

   where $\boldsymbol{X} = \mathbf{1}$ is an $(n \times 1)$ vector of 1's. So $\boldsymbol{X}^{\text{T}}\boldsymbol{X} = n$, $\boldsymbol{X}^{\text{T}}\boldsymbol{Y} = \sum Y_i$, which gives

   $$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\text{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\text{T}}\boldsymbol{Y} = \frac{1}{n}\sum Y_i = \bar{Y} = \widehat{\beta}_0,$$

   $$\text{var}[\widehat{\boldsymbol{\beta}}] = (\boldsymbol{X}^{\text{T}}\boldsymbol{X})^{-1}\sigma^2 = \frac{\sigma^2}{n}.$$

2. No-intercept model

   We saw that this example fits the General Linear Model with

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

   So $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \sum x_i^2$ and $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} = \sum x_i Y_i$, and we can calculate

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} = \frac{\sum x_i Y_i}{\sum x_i^2} = \widehat{\beta}_1,$$

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \frac{\sigma^2}{\sum x_i^2}.$$