NOTATION FOR MULTIPLE LINEAR REGRESSION

Response variable: Y (or y)

Predictor variables: X₁, X₂, … , Xₚ.

Note:
1. This is a change in notation: the subscript on the X’s now denotes a different variable, not a different observation.

2. p = number of predictor variables

So we would use x₁, x₂, … , xₚ to denote the values of X₁, X₂, … , Xₚ at one observation (i.e., for one case).

For short:

\[
X \quad \text{(or } X \text{ if handwritten)} = \begin{bmatrix} x₁ \\ x₂ \\ \vdots \\ xₚ \end{bmatrix} (or \begin{bmatrix} X₁ \\ X₂ \\ \vdots \\ Xₚ \end{bmatrix})
\]

(to refer to the random variables)

\[
x \quad \text{(or } \bar{x} \text{)} = \begin{bmatrix} x₁ \\ x₂ \\ \vdots \\ xₚ \end{bmatrix} (or \begin{bmatrix} x₁ \\ x₂ \\ \vdots \\ xₚ \end{bmatrix})
\]

(to refer to specific values of the r.v.'s)

Example:

E(Y|X) (or E(Y|\bar{x}) ) is short for

\[ E(Y | x₁, x₂, \ldots, xₚ) \]

\[ = E(Y | X₁ = x₁, X₂ = x₂, \ldots, Xₚ = xₚ) \]
To label data:

First observation: \( x_{11}, x_{12}, \ldots, x_{1p}, y_1 \)

Second observation: \( x_{21}, x_{22}, \ldots, x_{2p}, y_2 \)

\[ \vdots \]

\( n \)th observation: \( x_{n1}, x_{n2}, \ldots, x_{np}, y_n \)

Thus:

- \( n \) still = number of observations
- Subscript on \( y \) has same meaning as before (observation number)
- First subscript on \( x \) = observation number
- Second subscript on \( x \) = variable number
- i.e., \( x_{ij} \) = value of the \( j \)th predictor at the \( i \)th observation.

For short:

\[
\begin{bmatrix}
  x_{i1} \\
  x_{i2} \\
  \vdots \\
  x_{ip} \\
\end{bmatrix}
\]

\( \text{(or } \begin{bmatrix}
  x_{1} \\
  x_{2} \\
  \vdots \\
  x_{n} \\
\end{bmatrix} \text{)} \) -- the vector of
values of the predictor variables at observation \( i \).

The general goal of multiple regression:

Study how \( Y \mid x \) changes as \( x \) changes.

Example: Bic Mac

\( Y \) = the cost of a Big Mac in various countries
\( X_i \)'s = various economic indicators.

We'll use Bread, TeachSal, TeachTax, BusFare

Thus \( p = \_ \).