NOTATION FOR MULTIPLE LINEAR REGRESSION

Response variable Y (or y).
Predictor variables $X_1, X_2, \ldots, X_p$.

Note:

1. This is a change in notation: the subscript *on the X's* now denotes *a different variable, <u>not</u> a different observation.*
2. p = number of predictor variables

So we would use $x_1, x_2, \ldots, x_p$ to denote the values of $X_1, X_2, \ldots, X_p$ at *one* observation (i.e., for one case).

For short:

$$\mathbf{X} \text{ (or } \underline{X} \text{ when handwritten)} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{bmatrix} \text{ (or } \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{pmatrix})$$

(to refer to the random variables)

$$\mathbf{x} \text{ (or } \underline{x}) = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{bmatrix} \text{ (or } \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{pmatrix}) \qquad \text{(to refer to specific values of the r.v.'s)}$$

Example:

$E(Y|\mathbf{x})$ (or $E(Y|\underline{x})$) is short for $E(Y| x_1, x_2, \ldots, x_p)$
$= E(Y| X_1= x_1, X_2 =x_2, \ldots, X_p = x_p)$.

To label data:

First observation: $x_{11}, x_{12}, \ldots, x_{1p}, y_1$
Second observation: $x_{21}, x_{22}, \ldots, x_{2p}, y_2$
.
.
.
n^{th} observation: $x_{n1}, x_{n2}, \ldots, x_{np}, y_n$

Thus:

- n still = number of observations
- subscript on y is same as before
- first subscript on x is the observation number
- second subscript on x is the variable number
- i.e., x_{ij} = value of the j^{th} predictor at the i^{th} observation.

For short:

$$\mathbf{x}_i \text{ (or } \underline{x}_i) = \begin{bmatrix} x_{i1} \\ x_{i2} \\ . \\ . \\ . \\ x_{ip} \end{bmatrix} \text{ (or } \begin{pmatrix} x_{i1} \\ x_{i2} \\ . \\ . \\ . \\ x_{ip} \end{pmatrix} \text{) --}$$

the vector of values of the predictor variables at observation i.

The general goal of multiple regression:

Study how Y|**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, BusFar, so p = ___.