Multiple linear regression (MLR)

- **Multiple** linear regression: multiple **independent** variables ($\boldsymbol{X}$)
- **Multivariate** linear regression: multiple **dependent** variables ($\boldsymbol{Y}$)
- In real life applications, we are often interested in obtaining a relationship between X and Y **conditional** on other variables
- E.g. Modelling electricity cost and number of people in the household, conditional on whether there are solar panels installed

Setup

Consider data of the form

$$
\begin{gathered}
(y_1, x_{11}, x_{12}, \dots, x_{1r}) \\
(y_2, x_{21}, x_{22}, \dots, x_{2r}) \\
\vdots \\
(y_n, x_{n1}, x_{n2}, \dots, x_{nr})
\end{gathered}
$$

So we have n subjects with r predictors.

MLR model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_r x_{ir} + \epsilon_i,$$

where

$$\underline{\epsilon_i \sim i.i.d.\, N(0, \sigma^2),}$$

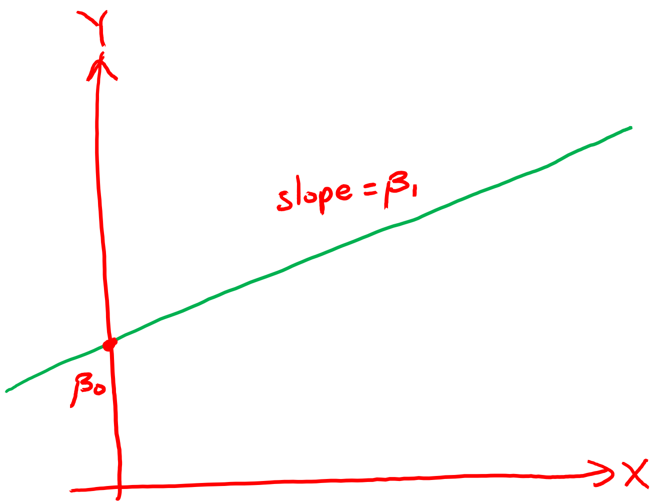
for $i = 1,2, \ldots, n$.

$$Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_r x_{ir}, \sigma^2)$$

independently for $i = 1, 2, \ldots, n$.

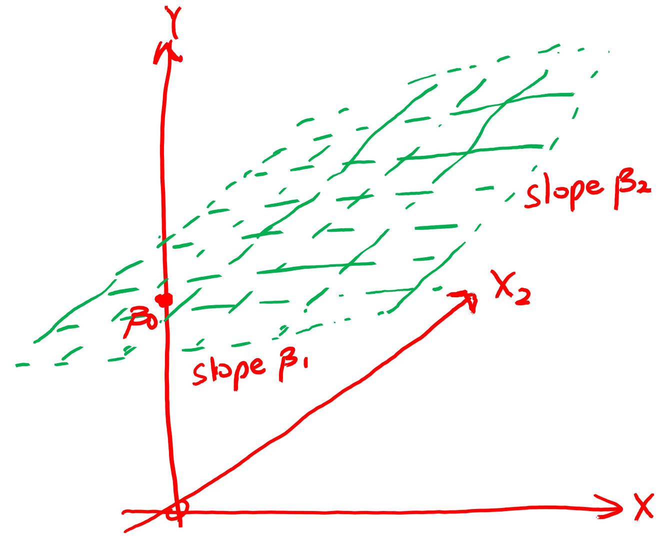
Simple linear regression (SLR)

(1 independent variable)



Multiple linear regression (MLR)

(2 independent variables)



Matrix formulation

Let

$$\underset{(n\times 1)}{\boldsymbol{Y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \underset{n\times(r+1)}{\boldsymbol{X}} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1r} \\ 1 & x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nr} \end{pmatrix}, \underset{(r+1)\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \boxed{\beta_0} \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix}, \text{ and } \underset{(n\times 1)}{\boldsymbol{\epsilon}} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

The multiple regression model can then be formulated as

$$\boxed{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.} \quad \text{where } \varepsilon \sim N_n(0, \sigma^2 I_n)$$