Multiple Regression

- Recall that the simple linear regression model can be expressed

\[ Y_i = \alpha + \beta X_i + \epsilon_i, \]

where \( \alpha \) and \( \beta \) are the unknown slope and intercept, and \( \epsilon \) is unknown random error with mean zero and constant variance for every observation.

The variable \( X_i \) is called the explanatory variable, independent variable, predictor, regressor, or covariate.

In simple linear regression, \( X_i \) is a scalar variable (a single number). In many cases, we may have \( d \geq 1 \) explanatory variables that are believed to influence the value of \( Y \). These explanatory variables can be stored in a column vector \( X_i = (X_{i1}, X_{i2}, \ldots, X_{id})' \).

The following linear model allows all \( d \) explanatory variables to influence the response:

\[ Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_d X_{id} + \epsilon_i. \]

We can express this model in matrix notation as

\[ Y_i = \alpha + \beta' X_i + \epsilon_i. \]