The **response variable** (weight) is plotted along the y-axis while the **explanatory variable** (height) is plotted along the x-axis. Deciding which variables are responses and which variables are explanatory factors is not always easy in interacting systems such as the climate. However, it is an important first step in formulating the problem in a testable (model-based) manner.

The cloud of points in a scatter plot can often (but not always!) be imagined to lie inside an ellipse oriented at a certain angle to the x-axis. Mathematically, the simplest description of the points is provided by the additive linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{6.1}$$

where $\{y_i\}$ are the values of the response variable, $\{x_i\}$ are the values of the explanatory variable, and $\{\epsilon_i\}$ are the left-over noisy **residuals** caused by **random effects** not explainable by the explanatory variable. It is normally assumed that the residuals $\{\epsilon_i\}$ are uncorrelated Gaussian noise, or to be more precise, a sample of independent and identically distributed (i.i.d.) normal variates.

The **model parameters** β_0 and β_1 are the y-intercept and the slope of the linear fit. They can be **estimated** using **least squares** by minimising the sum of squared residuals

$$SS = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{6.2}$$

By solving the two simultaneous equations

$$\frac{\partial SS}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{6.3}$$

$$\frac{\partial SS}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \tag{6.4}$$