

1 The linear model

The *linear model* is a theoretical framework that unifies a number of statistical concepts, like ANOVA and regression.

1.1 Least squares estimation

1.1.1 Statement of the problem

The general problem is this: We wish to model a linear relationship between a response variables Y and p predictor variables X_1, X_2, \dots, X_p . In other words:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (1)$$

Here, the β 's are the coefficients corresponding to the X 's. Now, assume that we have n data points, so that y_i corresponds to $x_{i1}, x_{i2}, \dots, x_{ip}$. In matrix form equation (1) now becomes:

$$y = X\beta \quad (2)$$

Here, $y \in \mathbb{R}^{n \times 1}$, $X \in \mathbb{R}^{n \times p}$ and $\beta \in \mathbb{R}^{p \times 1}$. Given y and X , we seek the best fit for β .