The General Linear Model

In a **general linear model**

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi} + \epsilon_i$$

the **response** $y_i, i = 1, \ldots, n$ is modelled by a linear function of **explanatory** variables $x_j, j = 1, \ldots, p$ plus an error term.

General and Linear

Here **general** refers to the dependence on potentially more than one explanatory variable, v.s. the **simple linear model**:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

The model is *linear in the parameters*, e.g.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_i$$
$$y_i = \beta_0 + \gamma_1 \delta_1 x_1 + \exp(\beta_2) x_2 + \epsilon_i$$

but not e.g.

$$y_i = \beta_0 + \beta_1 x_1^{\beta_2} + \epsilon_i$$
$$y_i = \beta_0 \exp(\beta_1 x_1) + \epsilon_i$$

Error structure

We assume that the errors ϵ_i are independent and identically distributed such that

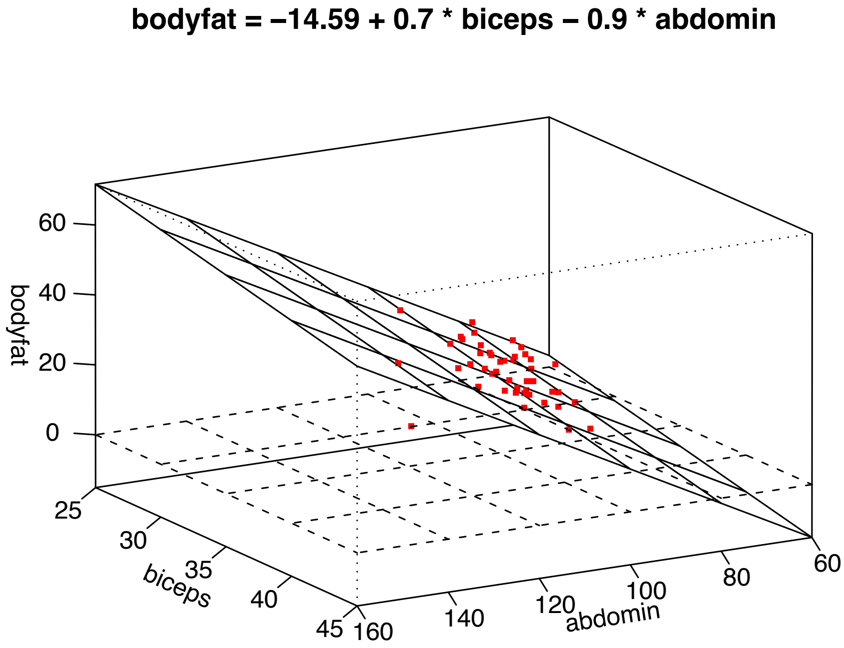
$$E[\epsilon_i] = 0$$
$$\text{and } \operatorname{var}[\epsilon_i] = \sigma^2$$

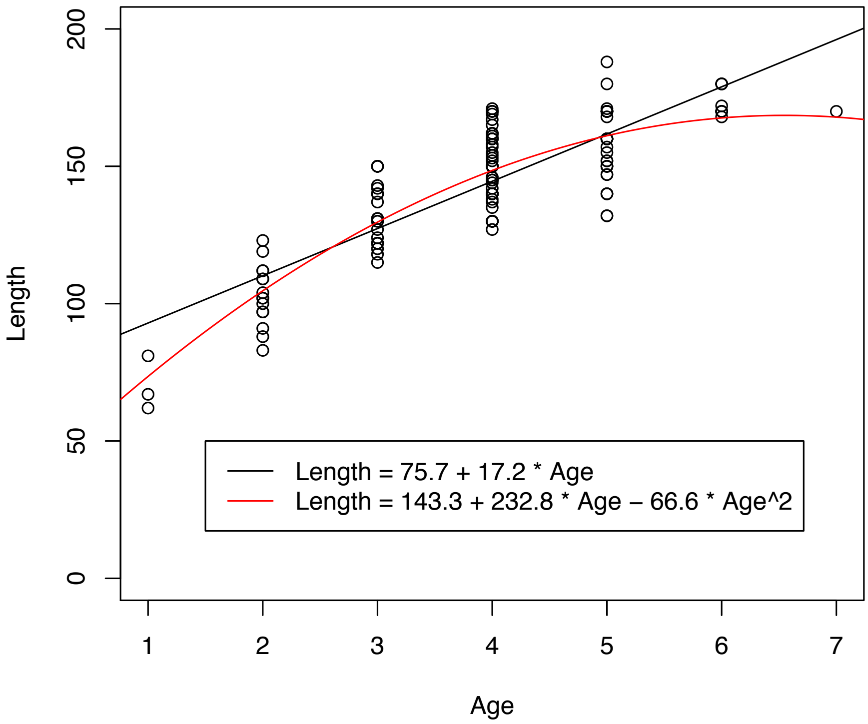
Typically we assume

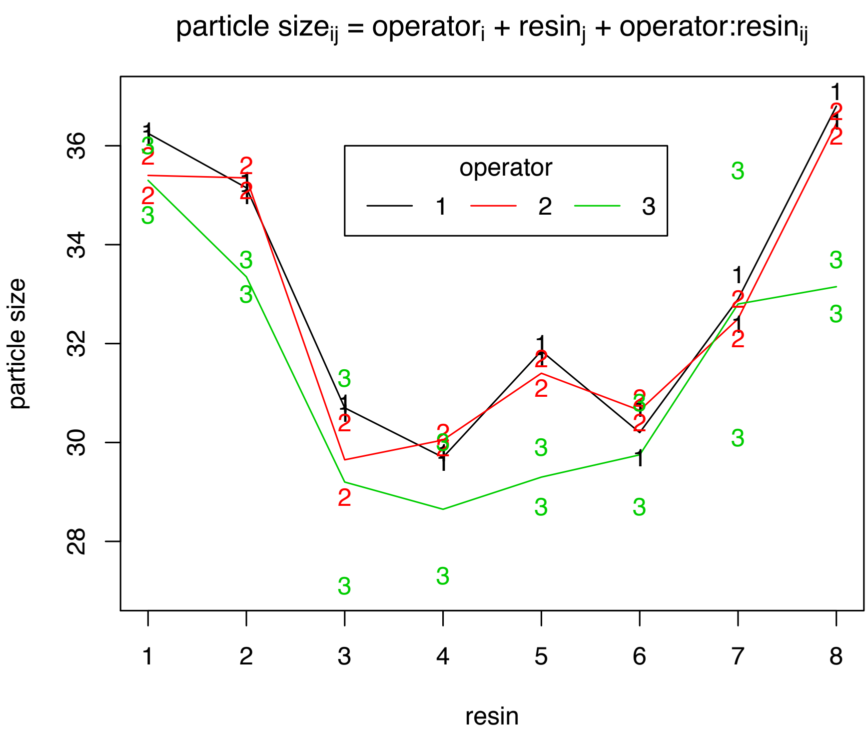
$$\epsilon_i \sim N(0, \sigma^2)$$

as a basis for inference, e.g. t-tests on parameters.

Some Examples




Length
0
50
100
150
200
1
2
3
4
5
6
7
Age
Length = 75.7 + 17.2 * Age
Length = 143.3 + 232.8 * Age – 66.6 * Age^2


particle size_ij = operator_i + resin_j + operator:resin_ij
operator
1
2
3
particle size
28
30
32
34
36
resin
1
2
3
4
5
6
7
8