Definition. *1. **Simple Linear Regression** A regression that is simple (1D) and linear in parameters (constant coefficients)*

2. Practically, in data set (x_i, y_i), we seek a fitted value for each x_i

$$\hat{y}_i = b_0 + b_1 x_i$$

and then setting $\hat{\beta}_0 = b_0$ and $\hat{\beta}_1 = b_1$.

residuals

$$\hat{e}_i = y_i - \hat{y}_i$$

Generally, we want to minimize, for some function g, the sum

$$\sum_{i=1}^{n} g(y_i - \hat{y}_i)$$

Consider

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^q$$

1. *$q = 0$ **0-1 loss***
2. *$q = 1$ **absolute loss***
3. *$q = 2$ **quadratic loss***
4. *$q = \infty$ susceptible to outliers*

***Quadratic loss** is the preferred loss function because*

1. ***Mean squared error (MSE)** is the most common way to measure error in statistics*
2. ***Gauss-Markov** theorem says that LSE when $q = 2$ has minimal variance*

Hence we pick b_0 b_1 that minimize sum of squares of residuals (RSS)

Definition. *Finding RSS*

$$\begin{aligned} RSS &= \sum_i \hat{e_i}^2 \\ &= \sum_i (y_i - \hat{y}_i)^2 \\ &= \sum_i (y_i b_0 - b_1 x_i)^2 \end{aligned}$$

To find b_0 and b_1 we take partial derivatives

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial RSS}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) x_i = 0$$

Set derivatives to zero yields ***normal equations***

$$\sum_{i=1}^{n} y_i = n b_0 + b_1 \sum_i x_i$$

$$\sum_i x_i y_i = b_0 \sum_i x_i + b_1 \sum_i x_i^2$$

Let $\overline{x} = \frac{1}{n} \sum_i x_i$ *and* $\overline{y} = \frac{1}{n} \sum_i y_i$*, rearrange normal equation*

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$b_1 = \frac{\sum_i x_i y_i - n \overline{x} \overline{y}}{\sum_i x_i^2 - n \overline{x}^2} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$