

# Retta di regressione

Retta  $y = a^* + b^*x$  dove

$$(a^*, b^*) = \arg \min_{a, b \in \mathbb{R}^2} \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Vale che:

$$b^* = \frac{\text{cov}(x, y)}{\text{var}(x)} \quad a^* = -b^*\bar{x} + \bar{y},$$

e:

$$\min_{a, b \in \mathbb{R}^2} \sum_{i=1}^n (y_i - a - bx_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 (1 - r(x, y)^2),$$

quindi:

- se  $|r(x, y)| \simeq 1$  i punti tendono ad essere allineati, se  $|r(x, y)| \simeq 0$  sono dispersi;
- $\text{sgn } r(x, y) = \text{sgn } b^*$  (perché i denominatori di  $r(x, y)$  e di  $b^*$  sono sempre positivi, e il numeratore è lo stesso).