Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) is a fundamental linear regression method that estimates the relationship between independent and dependent variables by minimizing the sum of squared differences (errors) between observed data points and predicted values. It calculates the "best-fitting" line, or hyperplane, by solving for parameters (slope and intercept) that minimize this error.

Key Aspects of OLS:

- **Goal:** To find the best-fitting straight line through data points, minimizing the sum of squared residuals.
- **Applications:** Widely used in econometrics, sociology, biology, and finance for forecasting and evaluating relationships.
- **Formula:** The slope (b_1) and intercept (b_0) are calculated by minimizing $\sum (y_i - y_i\hat{})^2$.
- **Assumptions:** Assumes a linear relationship, correct model specification, low multicollinearity, and that predictors are not correlated with the error term.
- **Limitations:** Sensitive to outliers, which can heavily influence the regression line.

Ordinary Least Squares (OLS) regression assumes a linear relationship between the dependent (target) variable and the independent (predictor) variables. The model aims to estimate the coefficients (also called betas) that provide the best fit to the data. The general formula for an OLS regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Where:

- Y is the dependent variable.
- β_0 is the intercept.
- β_j is the coefficient for each independent variable X_j.
- ϵ is the random error term.

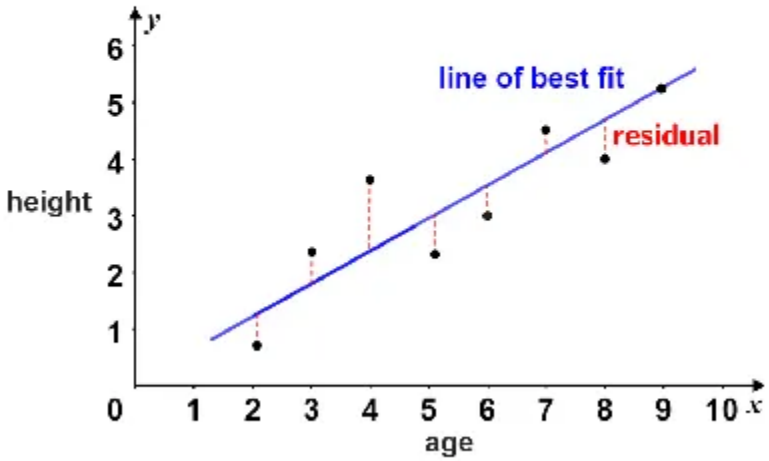
Ordinary Least Squares (OLS) is a cornerstone method in statistics and machine learning used for estimating the parameters of a linear regression model. It minimizes the sum of squared residuals (the differences between the observed and predicted values) to find the best-fitting line through a set of data points.

Press enter or click to view image in full size

OLS is a method for estimating the unknown parameters in a linear regression model. The goal is to find the line (or hyperplane in higher dimensions) that best fits the data by minimizing the sum of the squared differences between the observed values and the values predicted by the model.

The goal of OLS is to find the coefficient vector β that minimizes the sum of squared residuals (SSR). The residual for each observation is the difference between the observed value and the predicted value:

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



Predicting height with age