

## Tóm tắt lý thuyết Multiple Linear Regression (MLR)

Mô hình:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

trong đó

$y$  là biến phụ thuộc (response)

$X_1, X_2, \dots, X_k$  là biến độc lập (predictors)

$\beta_0$  là hệ số chặn (intercept)

$\beta_1, \dots, \beta_k$  là hệ số góc (slope) thể hiện tác động cận biên của từng biến  $X_j$  lên  $y$  khi các biến khác cố định.

$\varepsilon$  là sai số ngẫu nhiên (Gaussian,  $E[\varepsilon]=0$ ,  $\text{Var}(\varepsilon)=\sigma^2$ ).

Ước lượng OLS:

Tìm vectơ  $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$  sao cho tổng bình phương phần dư

$\sum_i (y_i - \beta_0 - \sum_j \beta_j X_{ij})^2$  nhỏ nhất.

Công thức:  $\beta = (X^T X)^{-1} X^T y$

Giả định chính:

- Tính tuyến tính giữa  $X$  và  $y$
- Sai số độc lập, phương sai không đổi (homoscedasticity)
- Không đa cộng tuyến mạnh giữa các  $X_j$
- Sai số tuân theo phân phối chuẩn

Đánh giá chất lượng:

- $R^2$ : tỉ lệ phương sai giải thích được
- Kiểm định t cho từng hệ số, kiểm định F cho toàn mô hình