

# Tóm Tắt: Các Mô Hình Hồi quy Cơ Bản

Trần Chí Vỹ

November 20, 2025

## Contents

<b>1</b>	<b>Chương 2: Hồi quy Tuyến tính (Linear Regression)</b>	<b>2</b>
1.1	Mục đích và Công thức Mô hình . . . . .	2
1.2	Phương pháp Xây dựng Mô hình . . . . .	2
1.3	Các hàm Excel chủ yếu . . . . .	2
1.4	Vai trò Chi tiết của Solver trong Hồi quy Tuyến tính . . . . .	2
<b>2</b>	<b>Chương 6: Hồi quy Logistic (Logistic Regression)</b>	<b>3</b>
2.1	Mục đích và Cơ chế Mô hình . . . . .	3
2.2	Phương pháp Xây dựng Mô hình . . . . .	3
2.3	Các hàm Excel . . . . .	3
2.4	Vai trò Chi tiết của Solver trong Hồi quy Logistic . . . . .	3

# 1 Chương 2: Hồi quy Tuyến tính (Linear Regression)

## 1.1 Mục đích và Công thức Mô hình

- Mục đích: Dự đoán **giá trị liên tục** ( $y$ ).

- Mô hình **Đa biến**:

$$y = m_1x_1 + m_2x_2 + \dots + m_nx_n + b$$

## 1.2 Phương pháp Xây dựng Mô hình

- Phương pháp:** Bình phương tối thiểu (Least Squares - OLS).
- Hàm Tối ưu (Loss Function):** Cực tiểu hóa Tổng Bình phương Phần dư (SSE).

$$\text{SSE} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

## 1.3 Các hàm Excel chủ yếu

Table 1: **Hàm Excel trong Hồi quy Tuyến tính**

Hàm Excel	Chức năng	Ví dụ
<b>LINEST</b>	Tính hệ số hồi quy đa biến	=LINEST(Y_range, X_range, TRUE, TRUE)
<b>SLOPE</b>	Hệ số góc ( $m$ ) (Đơn biến)	=SLOPE(Y_range, X_range)
<b>INTERCEPT</b>	Hệ số chặn ( $b$ ) (Đơn biến)	=INTERCEPT(Y_range, X_range)
<b>SUMPRODUCT</b>	Tính $\hat{y}$ (Phân tuyến tính)	=SUMPRODUCT(m, X) + b

## 1.4 Vai trò Chi tiết của Solver trong Hồi quy Tuyến tính

Solver là công cụ được sử dụng để tìm kiếm bộ hệ số  $m_i$  và  $b$  sao cho đường hồi quy phù hợp nhất với dữ liệu, đặc biệt hữu ích khi không sử dụng hàm LINEST hoặc khi có ràng buộc phức tạp.

- Mục tiêu Chính (Set Objective):** Solver được thiết lập để tìm giá trị **nhỏ nhất** (Min) của ô chứa công thức tính **Tổng Bình phương Phần dư (SSE)**.
- Biến Thay đổi (By Changing Variable Cells):** Các ô chứa giá trị ban đầu (thường là 0) của tất cả các hệ số  $m_1, m_2, \dots, m_n$  và hệ số chặn  $b$ .
- Cơ chế hoạt động:** Solver sử dụng các thuật toán tối ưu hóa (thường là **GRG Nonlinear** trong Excel) để lặp đi lặp lại, điều chỉnh giá trị của các hệ số  $m_i, b$  cho đến khi SSE đạt giá trị cực tiểu. Bộ hệ số cuối cùng là bộ hệ số tối ưu của mô hình OLS.

## 2 Chương 6: Hồi quy Logistic (Logistic Regression)

### 2.1 Mục đích và Cơ chế Mô hình

- Mục đích:** Dự đoán xác suất sự kiện xảy ra ( $P$ ) và phân loại (0/1).
- Hàm Sigmoid/Logistic:**

$$P = \frac{1}{1 + e^{-z}}$$

### 2.2 Phương pháp Xây dựng Mô hình

- Phương pháp:** Ước lượng Khả năng Hợp lý Tối đa (Maximum Likelihood Estimation - MLE).
- Hàm Tối ưu (Loss Function):** Cực tiểu hóa Log Loss.

$$\text{Log Loss} = - \sum_{i=1}^N [y_i \ln(P_i) + (1 - y_i) \ln(1 - P_i)]$$

### 2.3 Các hàm Excel

Table 2: **Hàm Excel trong Hồi quy Logistic**

Hàm Excel	Chức năng	Ví dụ
SUMPRODUCT	Tính phần tuyến tính ( $z$ )	=SUMPRODUCT(m, X) + b
EXP	Tính $e^x$ , cần thiết cho Sigmoid	=1/(1+EXP(-z))
LN	Log tự nhiên, dùng để tính Log Loss	-LN(P)
IF	Tính Log Loss từng dòng	=IF(y=1, -LN(P), -LN(1-P))

### 2.4 Vai trò Chi tiết của Solver trong Hồi quy Logistic

Trong Hồi quy Logistic, Solver là công cụ \*\*bắt buộc\*\* để tìm ra các hệ số tối ưu, vì không có công thức giải tích đóng (closed-form solution) như OLS. Phương pháp được sử dụng là MLE, và Solver giúp tối ưu hóa nó.

- Mục tiêu Chính (Set Objective):** Tối ưu hóa hàm Khả năng Hợp lý (Likelihood). Có hai cách thiết lập tương đương:
  - Đặt mục tiêu là **nhỏ nhất** (Min) của ô chứa công thức tính **Tổng Log Loss** (Hàm lỗi).
  - Đặt mục tiêu là **lớn nhất** (Max) của ô chứa công thức tính **Tổng Log Likelihood** (Độ phù hợp).
- Biến Thay đổi (By Changing Variable Cells):** Các ô chứa giá trị của tất cả các hệ số  $m_i$  và hệ số chặn  $b$ .

- **Cơ chế hoạt động:** Solver lặp đi lặp lại (dùng thuật toán như Newton-Raphson hoặc Gradient Descent thông qua GRG Nonlinear) để tìm ra bộ hệ số  $m_i, b$  làm cho xác suất quan sát dữ liệu thực tế là cao nhất (tối đa hóa Likelihood) hay tương đương là tối thiểu hóa Log Loss.