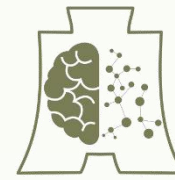


# Multiple Linear Regression Introduction

National Kaohsiung University of Sciences and Technology  
Department of Finance and Information, Professor  
AI Fintech Center, Director  
Lin, Ping-Chen



- 什麼是複迴歸(多元線性迴歸)?
- 複迴歸的公式與計算
- 複迴歸的假設
- 複迴歸在金融中的應用



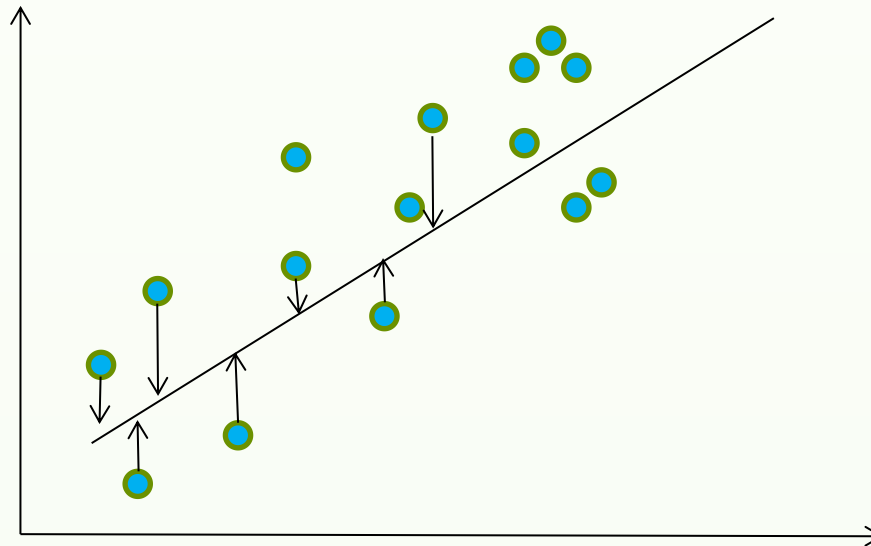
# Multiple Linear Regression Introduction

- 複回歸是一種統計技術，使用多個自變數來預測依變數的結果
- MLR的目標是建模解釋自變數和依變數之間的線性關係
- 它是最小平方法（OLS）迴歸的擴展，因為它涉及多於一個解釋變數。
- MLR在計量經濟學和財務推論中廣泛應用。
- 複迴歸用於做預測、解釋財務變數之間的關係和檢驗現有理論。

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$y_i$  : actual value

$\hat{y}_i$  : forecast value



- MLR 是一種統計技術，它使用多個自變數來預測依變數的結果

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon$$

其中，對於  $i = n$  個觀測值：

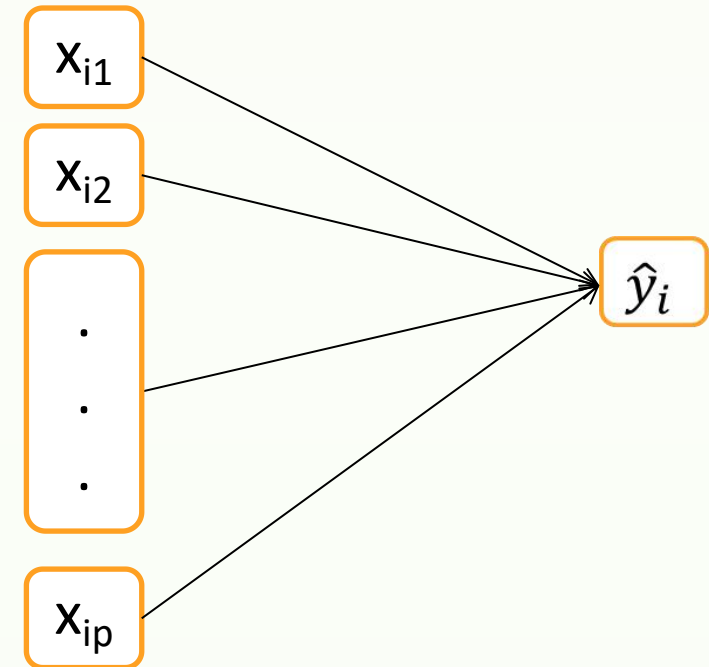
$\hat{y}_i$  = 依變數

$x_i$  = 解釋變數

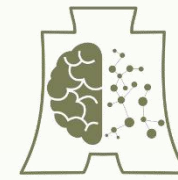
$\beta_0$  = 常數項

$\beta_p$  = 每個解釋變數的斜率係數

$\varepsilon$  = 模型的誤差項

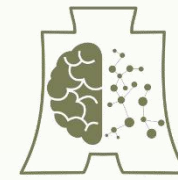


## 複迴歸的假設



- 自變數和依變數之間必須是線性關係
- 自變數之間不應該高度相關
- $y_i$  觀察值應從樣本中獨立隨機選取
- 殘差應該呈正態分佈，均值为0，方差為  $\sigma^2$
- 決定係數是一種統計指標，用於衡量結果的變化有多少可以透過自變數的變化來解釋。
- 隨著更多變數加入 MLR 模型中， $R^2$  始終會增加，即使自變數可能與依變數無關。
- $R^2$  只能介於 0 和 1 之間，其中 0 表示任何自變數都無法預測結果，1 表示可以從自變數預測結果而不會出現錯誤。
- 當所有變數保持不變（其他條件相同）時， $\beta$  係數有效。

## 複迴歸在金融中的應用



AI.FINTECH

AI 金融科技中心

- 任何包含多個變數的計量經濟模型可能都是多元的
- 因子模型比較兩個或多個因子來分析變數之間的關係及其結果表現
- Fama和French三因子模型擴展了資本資產定價模型（CAPM），在CAPM的市場風險因子中加入了規模風險和價值風險因子（這本身也是一個迴歸模型）
- MLR 是一種統計工具，用於根據兩個或多個自變數來預測依變數的結果。
- 如果只有一個自變數影響依變數，則簡單線性迴歸模型就足夠了
- 如果有多個變數影響，則需要複迴歸。



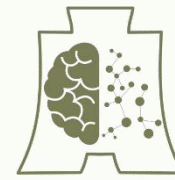
- <https://medium.com/@jason8410271027/%E5%AD%B8%E7%BF%92%E7%AD%86%E8%A8%98-%E7%B7%9A%E6%80%A7%E5%9B%9E%E6%AD%B8-linear-regression-38b17484ee0a>
- <https://medium.com/@jason8410271027/%E5%AD%B8%E7%BF%92%E7%AD%86%E8%A8%98-%E5%A4%9A%E5%85%83%E7%B7%9A%E6%80%A7%E5%9B%9E%E6%AD%B8-multiple-linear-regression-%E7%90%86%E8%AB%96%E7%AF%87-bb09cf72af39>
- <https://python.plainenglish.io/understanding-multiple-linear-regression-in-machine-learning-58e981ce7747>
- <https://medium.com/@chih.sheng.huang821/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E5%9F%BA%E7%A4%8E%E6%95%B8%E5%AD%B8-%E4%BA%8C-%E6%A2%AF%E5%BA%A6%E4%B8%8B%E9%99%8D%E6%B3%95-gradient-descent-406e1fd001f>





**Thank you.**

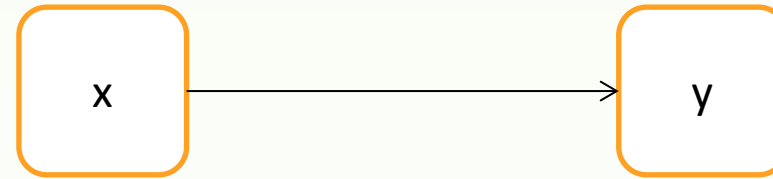




## 簡單線性迴歸介紹

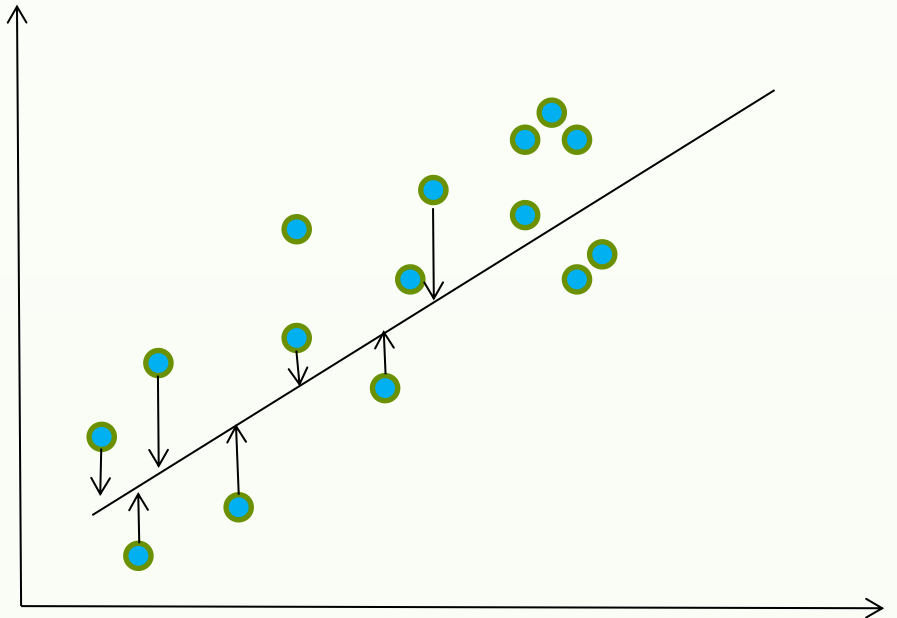
- 簡單線性迴歸:從資料中找出一條線, 可以透過這條線了解**自變數(x)**跟**應變數(y)**的關係
- $\beta_0$ 為截距項、 $\beta_1$ 為斜率、 $\epsilon$ 為誤差項

$$y = \beta_0 + \beta_1 x + \epsilon$$

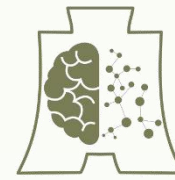


使用最小平方法估計參數:找出誤差最小的線

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

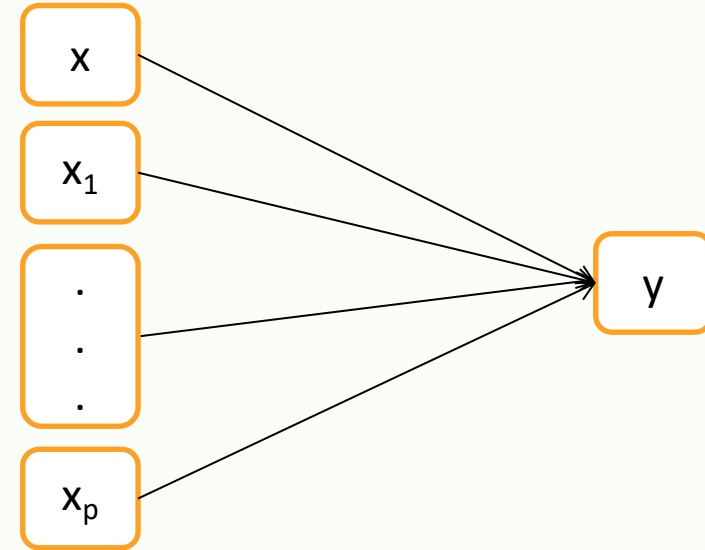


## 復迴歸介紹



- 從資料中找出一條線，可以透過這條線了解**多個自變數( $x_1 \sim x_p$ )**跟**應變數( $y$ )**的關係
- $\beta_0$ 為截距項、 $\beta_1 \sim \beta_p$ 為斜率、 $\epsilon$ 為誤差項

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \cdots + \beta_p x_p + \epsilon$$



- 復迴歸假設
  - 1.應變數和自變數之間的關係必須是線性的。
  - 2.自變數必須相互獨立。
  - 3.誤差的變異數在自變數的所有水準上應該是恆定的。
  - 4.誤差應呈常態分佈。
  - 5.自變數不應彼此高度相關。