

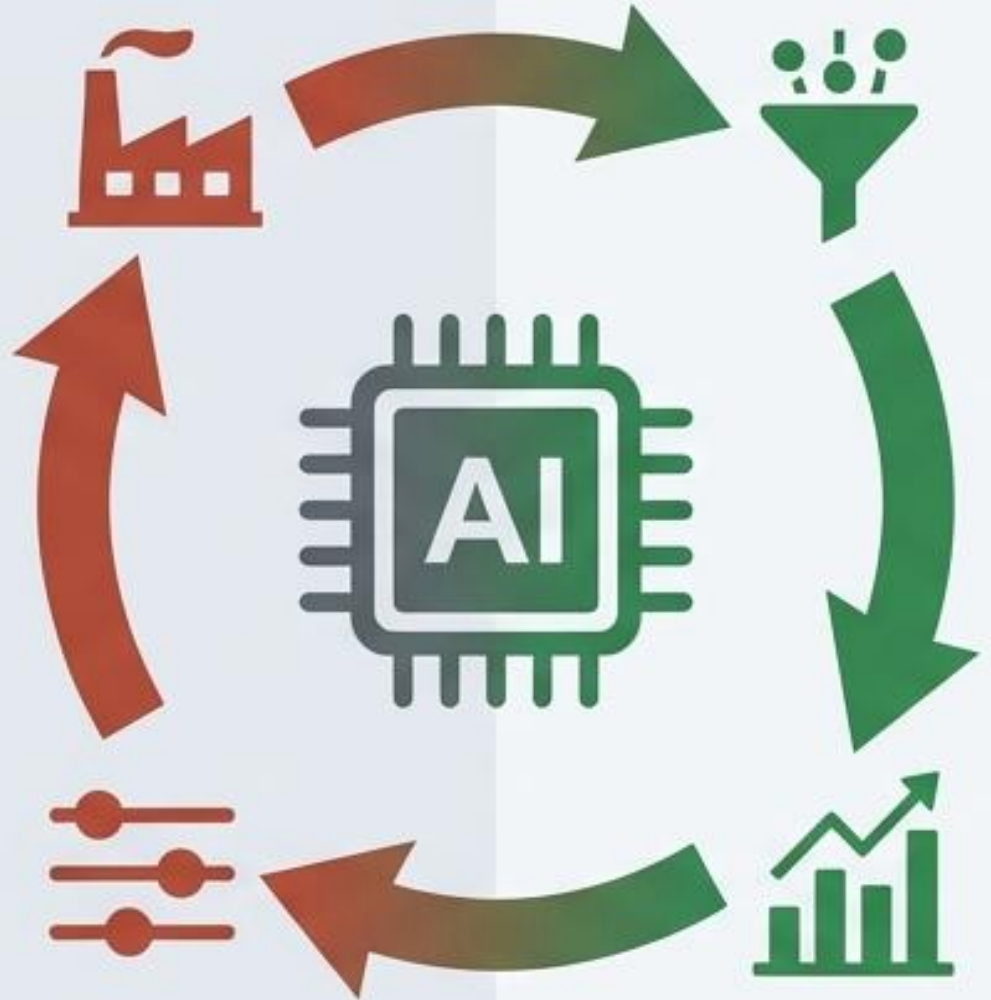
設備健康預診 **×** AI 數據診斷

講師：徐睿朋

AI智慧化解決方案：從隱性痛點到全面優勢

機械設備業痛點

-  資訊不足，數據零散
-  維護成本高，停機損失
-  無法預測，決策困難
-  參數固定，效率受限



AI 賦能的四大優勢

-  **即時監測**
掌握動態設備狀態
-  **預測維護**
降低停機與維護成本
-  **智慧決策**
AI統計分析優化運轉參數
-  **效率與品質提升**
節能、提升穩定性與良率



AI 實戰與原理解析：資料與統計

Understanding Data Before Machine Learning

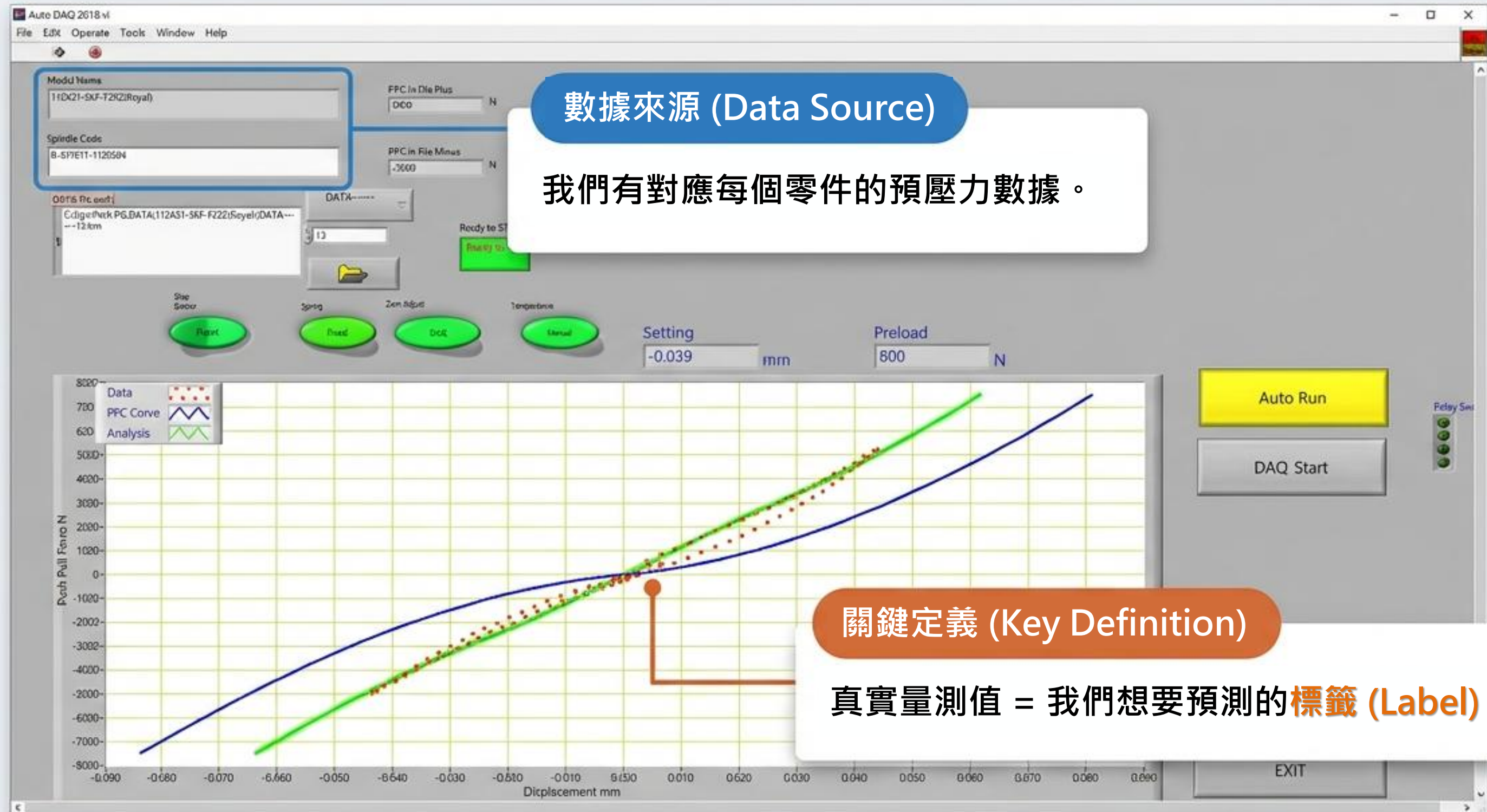
數據的迷霧：光看這些數字， 你能判斷軸承的預壓力嗎？

rot_amp_x	rot_amp_z	FTF_amp_x	FTF_amp_z	BSF_amp_x	BSF_amp_z	BPF0_amp_x	BPF0_amp_z	BPMI_amp_x	BPMI_amp_z
0.029373	0.028651	0.005189	0.003923	0.002446	0.001226	0.037831	0.0006666	0.018566	0.003494
0.029433	0.028695	0.005232	0.003979	0.001899	0.000929	0.036217	0.0006612	0.023625	0.00335
0.029387	0.028636	0.005302	0.004065	0.002355	0.001097	0.037953	0.0007601	0.02954	0.003181
0.029373	0.028651	0.005303	0.004065	0.002355	0.001097	0.037953	0.0007601	0.029554	0.003181
0.029433	0.028695	0.005329	0.004063	0.002346	0.001896	0.037953	0.0077616	0.029581	0.003181
0.029433	0.028695	0.005332	0.004065	0.002355	0.001097	0.037953	0.0077601	0.023625	0.003494
0.029433	0.028698	0.005337	0.004065	0.002355	0.001096	0.037382	0.0077616	0.023625	0.003494
0.029433	0.028698	0.005310	0.004065	0.002355	0.001097	0.037031	0.0077618	0.029544	0.003181
0.029433	0.028694	0.005331	0.004065	0.002355	0.001097	0.037953	0.0078637	0.029566	0.003368
0.029433	0.028698	0.005209	0.004065	0.002355	0.001097	0.037953	0.0077638	0.023525	0.003170
0.029373	0.028651	0.005221	0.004923	0.002446	0.001226	0.037831	0.0077666	0.023566	0.003494
0.029433	0.028695	0.005232	0.003923	0.002446	0.001226	0.037817	0.0076612	0.023624	0.003494
0.029387	0.028699	0.005332	0.004979	0.002459	0.001029	0.037217	0.0070702	0.023624	0.003181
0.029387	0.028636	0.005302	0.004065	0.002456	0.001226	0.037217	0.0077601	0.023625	0.003181
0.029387	0.028637	0.005302	0.004065	0.002355	0.001097	0.037953	0.0077601	0.029554	0.003181
0.029383	0.028636	0.005313	0.004065	0.002355	0.001097	0.037953	0.0077601	0.029554	0.003181
0.029383	0.028651	0.005322	0.004065	0.002355	0.001097	0.037953	0.0077681	0.022954	0.003181
0.029373	0.028651	0.005189	0.003923	0.002446	0.001226	0.037831	0.00666666	0.018566	0.003494
0.029433	0.028651	0.005189	0.003923	0.002446	0.001226	0.037817	0.00766666	0.018566	0.003494
0.029433	0.028695	0.005232	0.003979	0.001899	0.000929	0.036217	0.00706612	0.023625	0.00335
0.029387	0.028636	0.005302	0.004065	0.002355	0.001097	0.037953	0.00707601	0.029954	0.003181
0.029433	0.028651	0.005189	0.004923	0.002436	0.001226	0.036783	0.00766601	0.018566	0.003491
0.029433	0.028695	0.005232	0.004979	0.002399	0.001029	0.036217	0.00706612	0.023625	0.003433
0.029387	0.028636	0.005302	0.004065	0.002355	0.001097	0.036217	0.00767618	0.023625	0.003495
0.029387	0.028636	0.005302	0.004065	0.002355	0.001097	0.037953	0.00767611	0.029954	0.003181

痛點 (Pain Point)

單純看表格難以看出規律。
這就是為甚麼可視化與資料分析很重要。

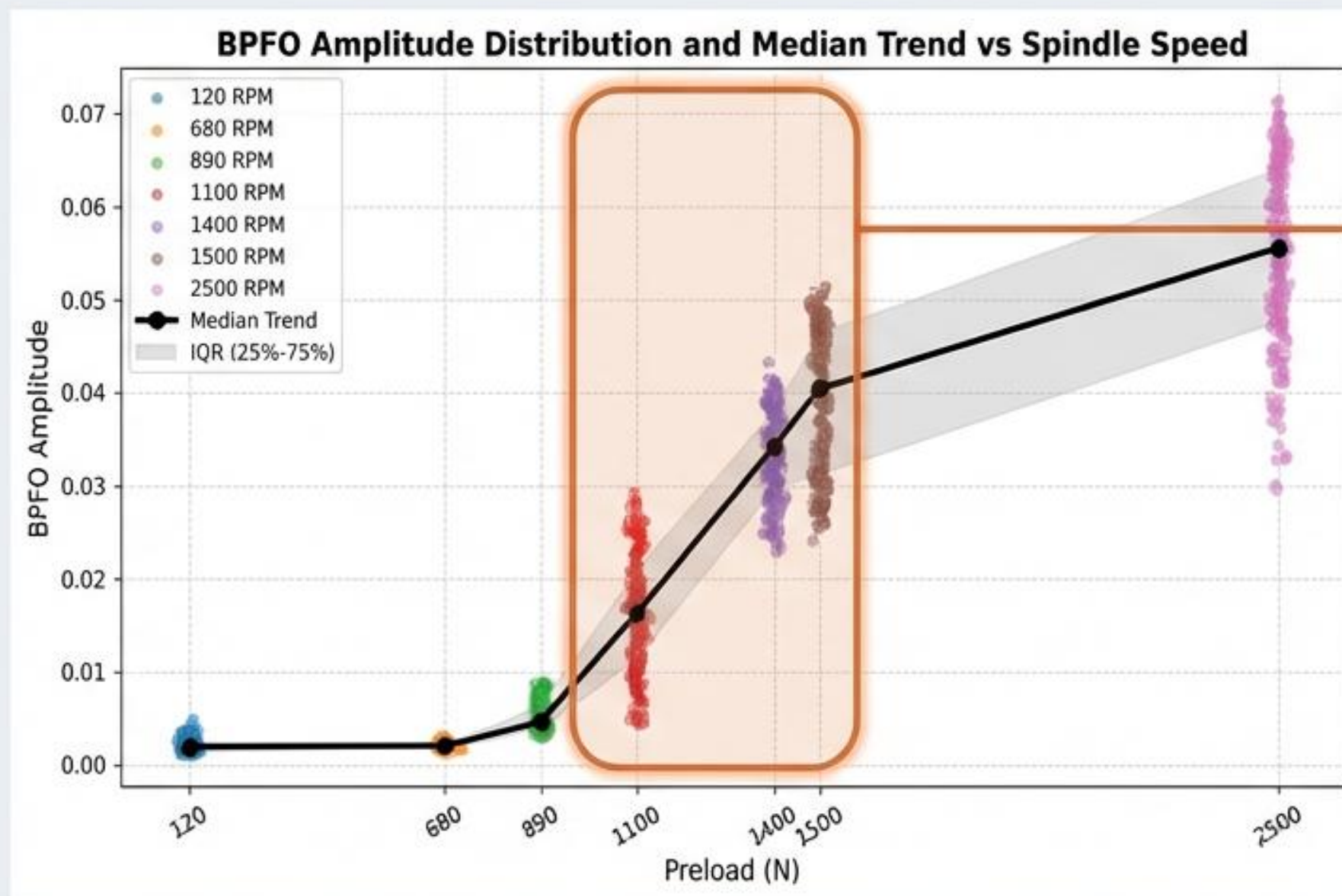
步驟一：賦予數據意義，建立 AI 模型的「學習目標」



步驟二：讓規律現身，從視覺化提取「學習特徵」

視覺洞察 (Visual Insight)

觀察不同預壓力
對應的不同振動
範圍。

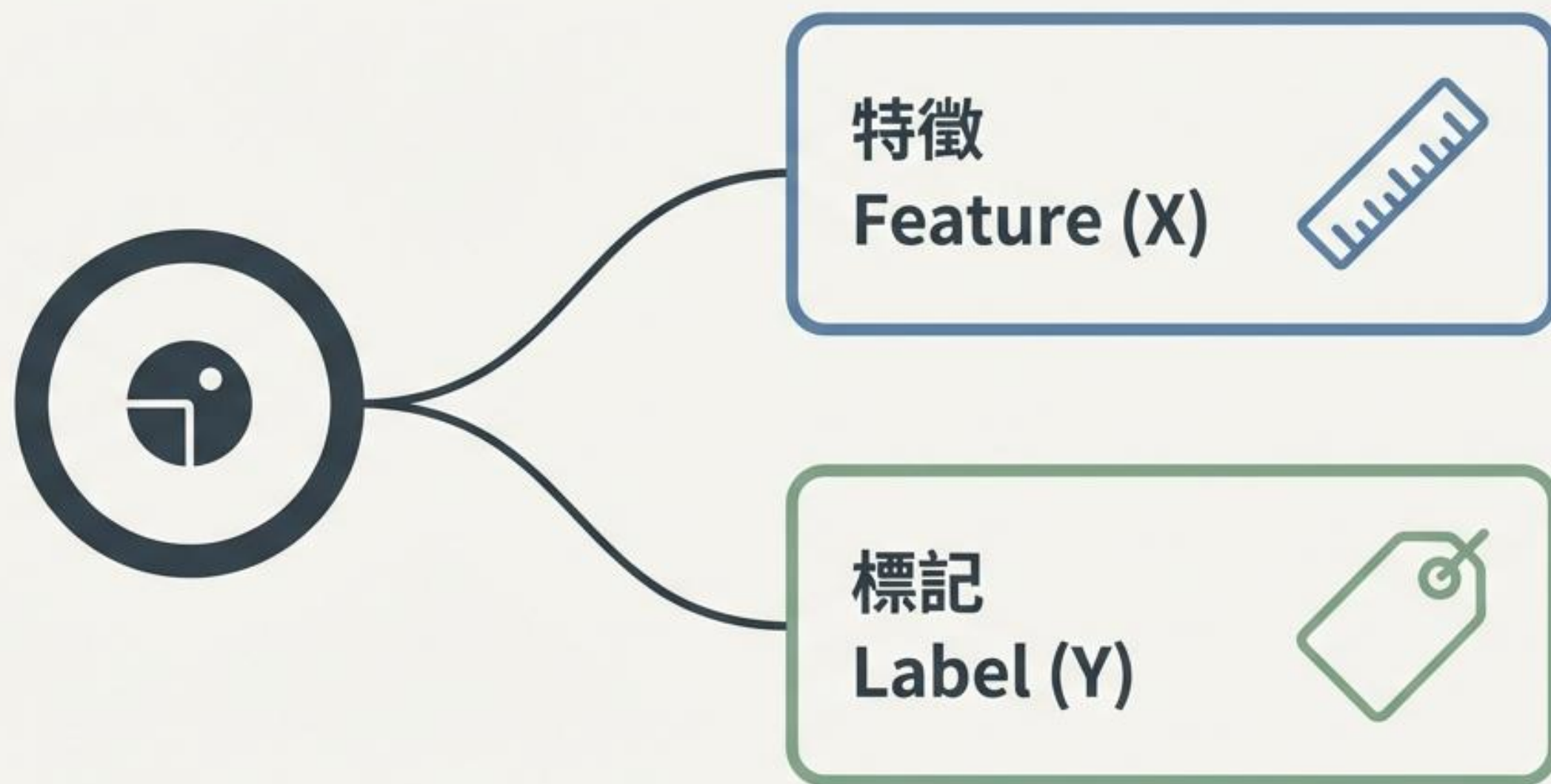


關鍵轉換 (Key Transition)

資料呈現群聚 →
可作為模型學習
特徵 (Feature)。

認識資料的基本結構

每一筆觀測資料都由兩個核心要素組成：



【教學範例：預測連續數值】 加州房價數據

➔ 任務類型：迴歸問題 (Regression)

➔ 本單元重點：透過統計圖表建立數據的「量感」與誤差觀念。

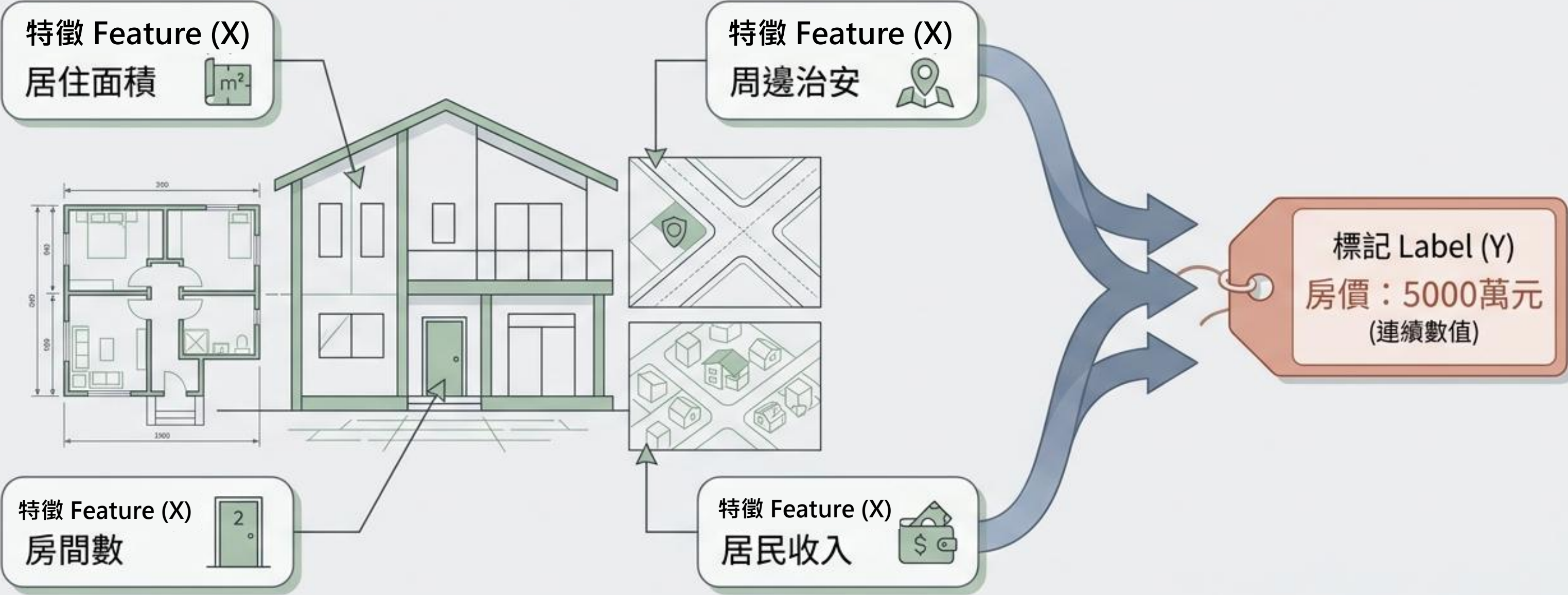


[房屋特徵] → 機器學習模型 → 預測房價金額：\$1020 萬
(與實際 1000 萬比對：誤差 20 萬)

「這任務不是看對錯，而是看誤差有多大。」

每一筆觀測資料的核心要素

模型透過學習多個「特徵 (X)」的組合，來預測出最終的「標記 (Y)」——在這個案例中，這是一個連續的數值（房價，單位：萬元）。

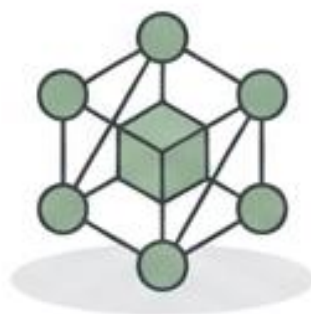


經典範例：加州房價資料集概況



20,640 筆資料

真實反映現實世界的
豐富度與雜訊。



8 個數值特徵

包含中間收入、屋齡、房
間數等關鍵屬性。



預測目標 (Y)

連續型數值（房價）。這
正是典型的「迴歸問題
(Regression)」核心。

敘述統計：建立對數據的「量感」

中間收入 (MedInc)



房屋年齡 (HouseAge)

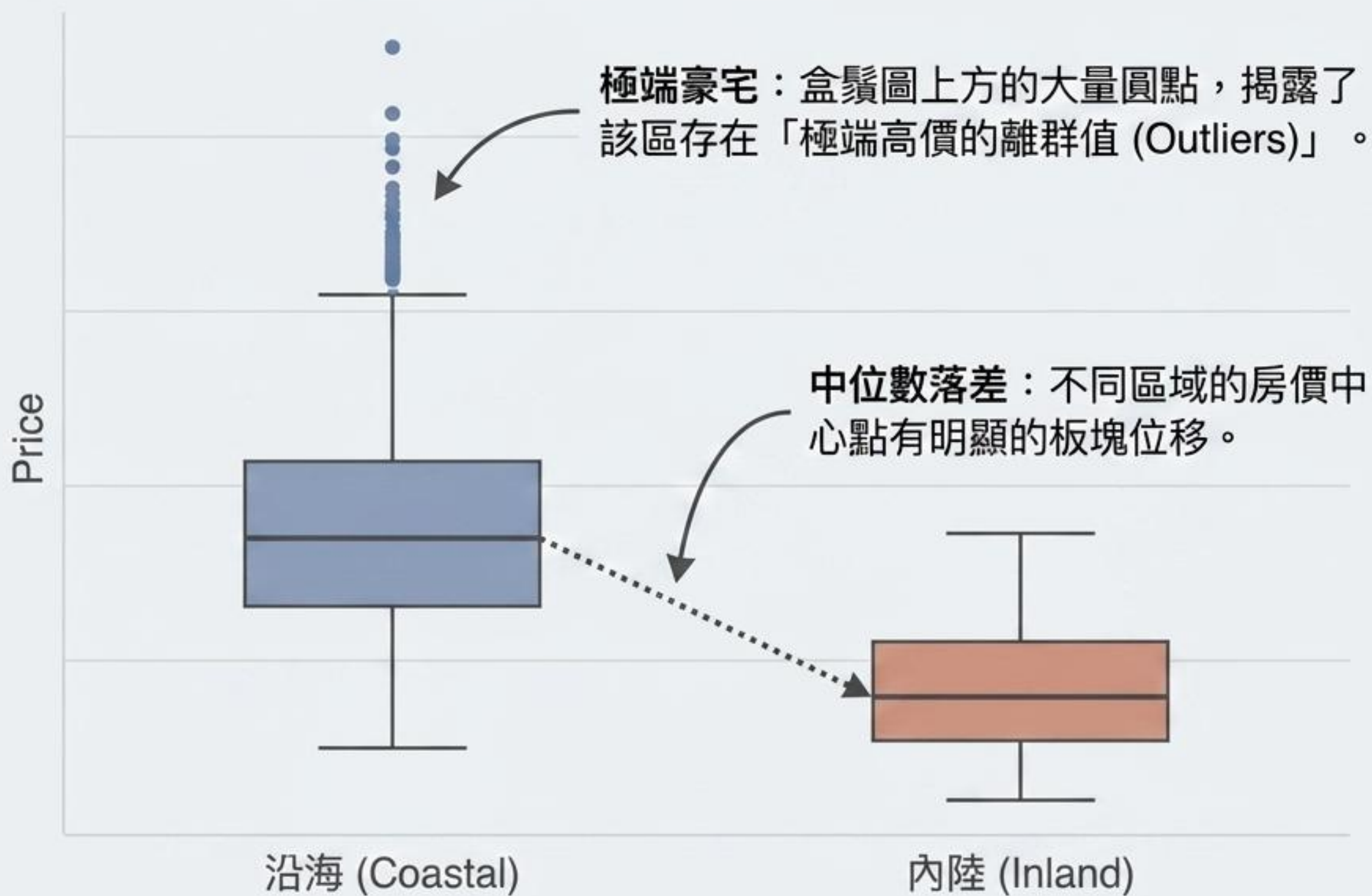


房價 (Price)

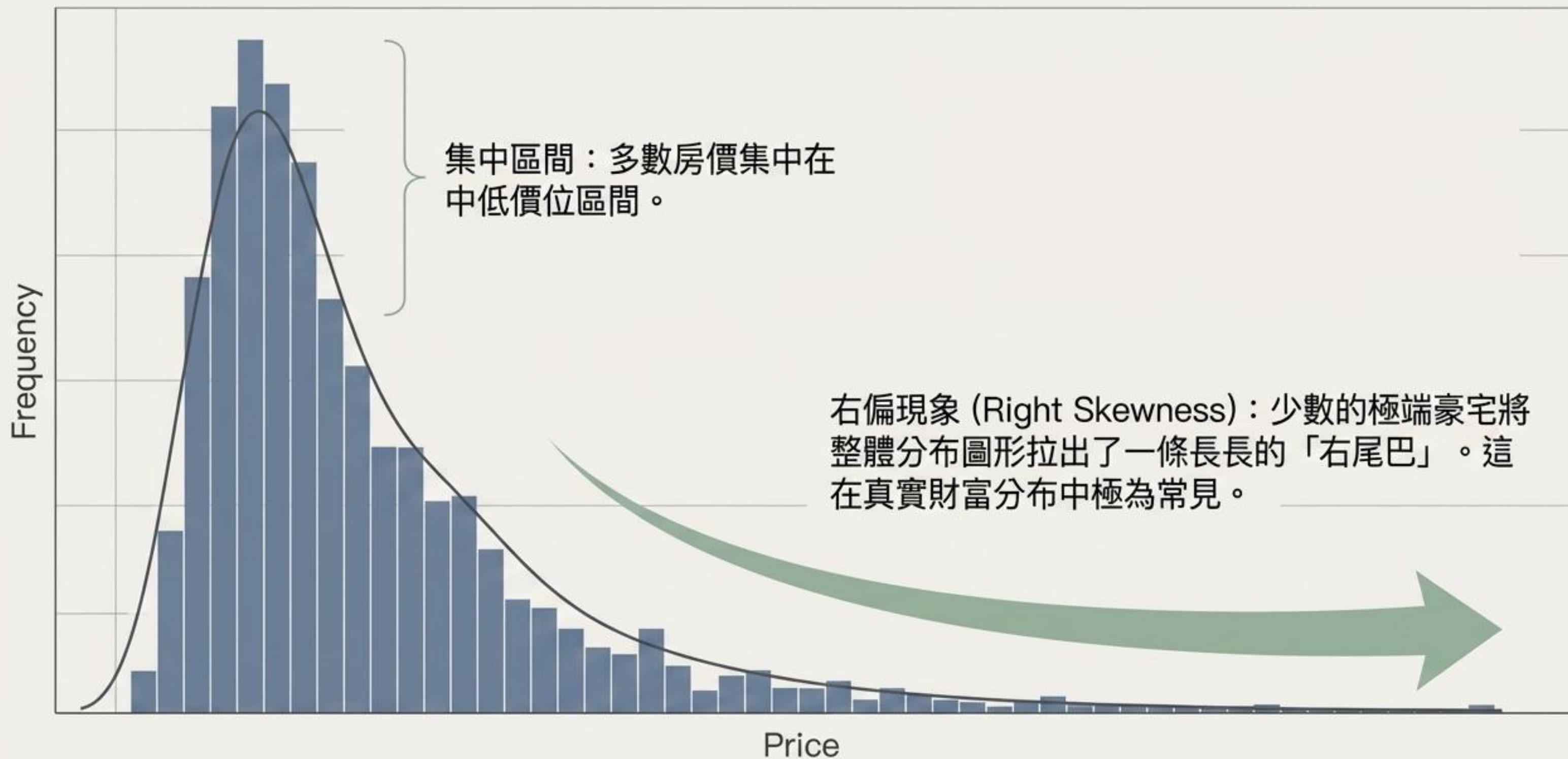


觀察重點：注意最大值 (Max) 與最小值 (Min) 之間的巨大差距。這能幫助我們快速掌握房價與收入在現實中的極端分布範圍。

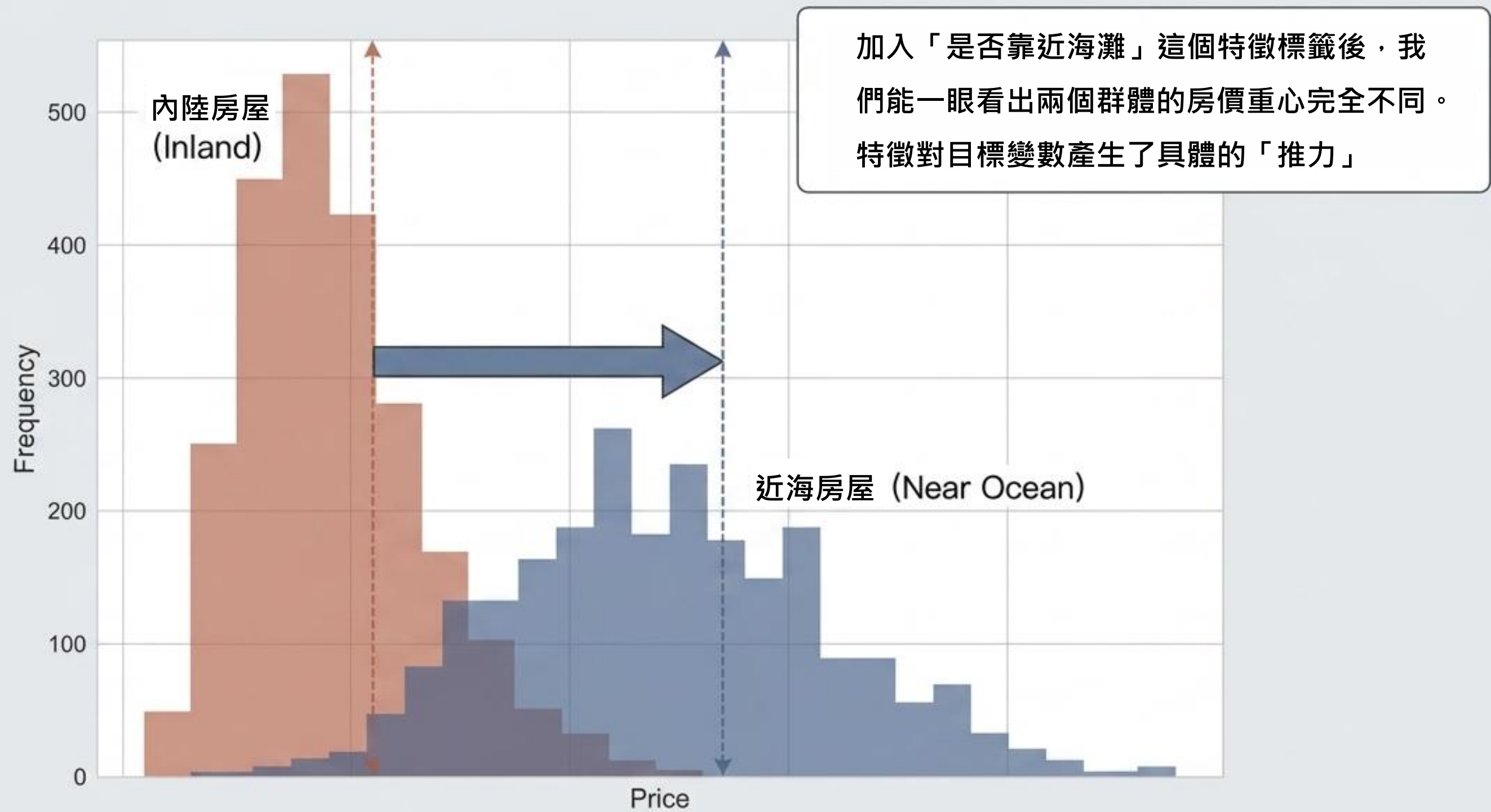
盒鬚圖 (Box Plot)：觀察分布與離群值



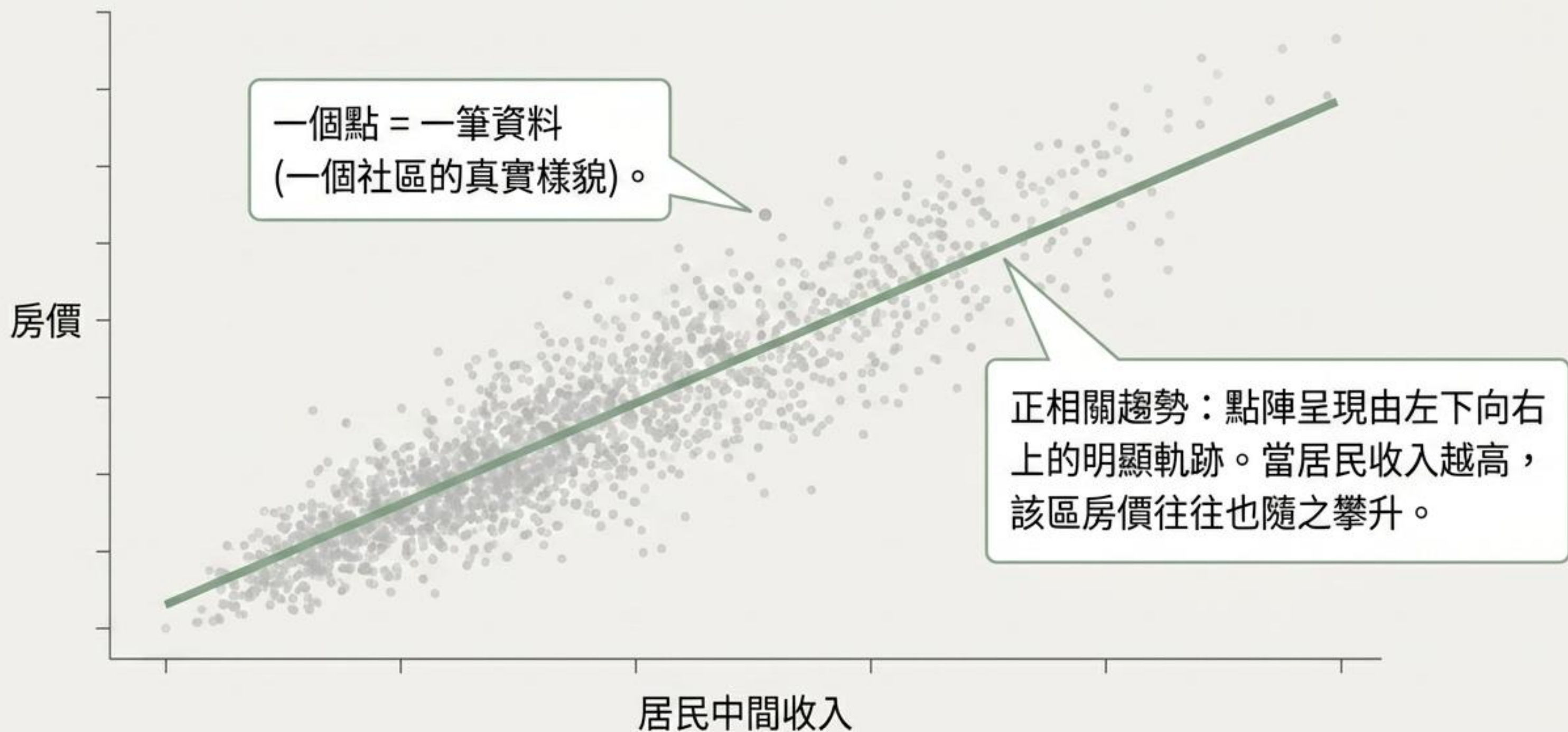
直方圖 (Histogram)：看見數值的機率形狀



特徵分組：揭露隱藏的房價引力



散佈圖 (Scatter Plot)：尋找連動關係



【基礎統計觀念】 什麼是離群值 (Outliers) ？



- **離群值定義**：在一個數據集中，顯著偏離其他絕大部分觀測值的極端數據點。
- **統計學特性**：離群值的存在會對傳統統計量（如平均值、標準差）產生劇烈的拉扯與影響。
- **本單元引導重點**：在進入任何數據分類或模型訓練之前，必須先透過統計圖表（如盒鬚圖）學會識別並找出這些分佈異常的數字。

【原理講解】連續型問題的評估指標

真實房價 —— 預測房價 —— 差距 (誤差)

1000 萬 —— 1020 萬 —— 20 萬

500 萬 —— 490 萬 —— 10 萬

2000 萬 —— 1970 萬 —— 30 萬

統計與評估這些差距的方法：

方法 1 (MAE)：直接算平均差距。

方法 2 (RMSE)：放大並懲罰大差距。

任務類型定義：屬於典型的「迴歸問題 (Regression)」。

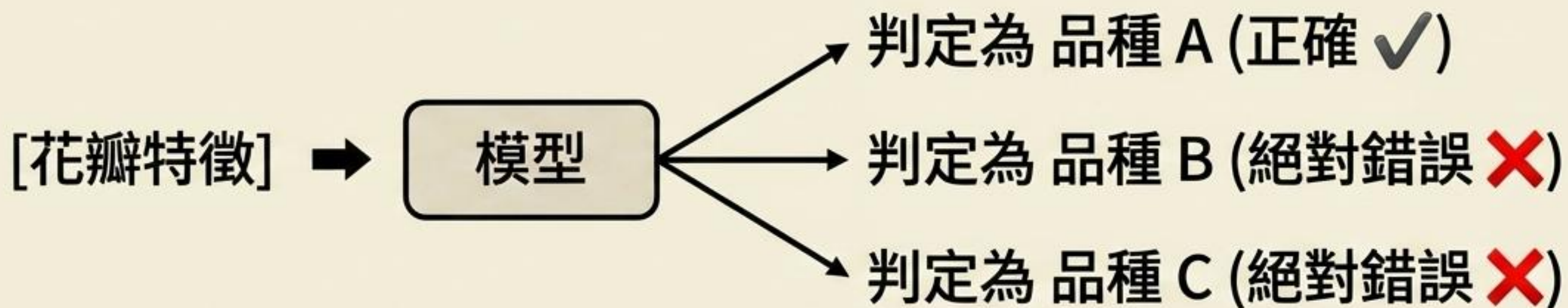
核心預測邏輯：此類任務之重點不在於判定絕對的對錯，而是評估「誤差的大小」。

本單元導引重點：透過統計圖表建立對數據的分佈量感，並理解如何統計預測與真實之間的差距總帳。

【教學範例：預測離散類別】

鳶尾花品種分類

- ➡ 任務類型：分類問題 (Classification)
- ➡ 本單元重點：略過重複的圖表，直奔「類別不平衡」的現實挑戰。



「這任務是非黑即白，對就是對、錯就是錯。」

經典範例：Iris Dataset 鳶尾花資料集



Petal (花瓣) Sepal (花萼)



Petal (花瓣) Sepal (花萼)



Petal (花瓣) Sepal (花萼)

共 150 筆資料
(乾淨、無缺失值)

分為 3 種鳶尾花
(Setosa, Versicolor, Virginica)

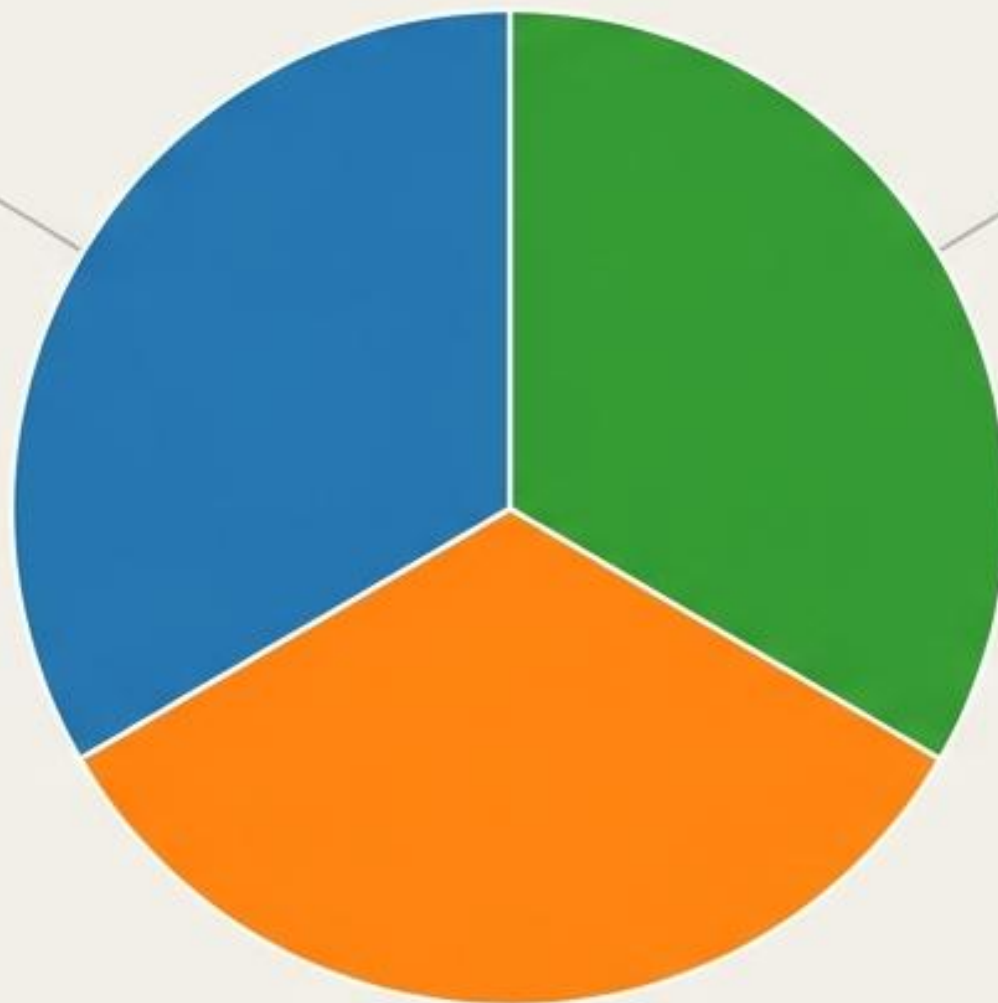
包含 4 個數值特徵
(長度與寬度，單位：cm)

類別數量平衡

理想世界：完美的類別平衡

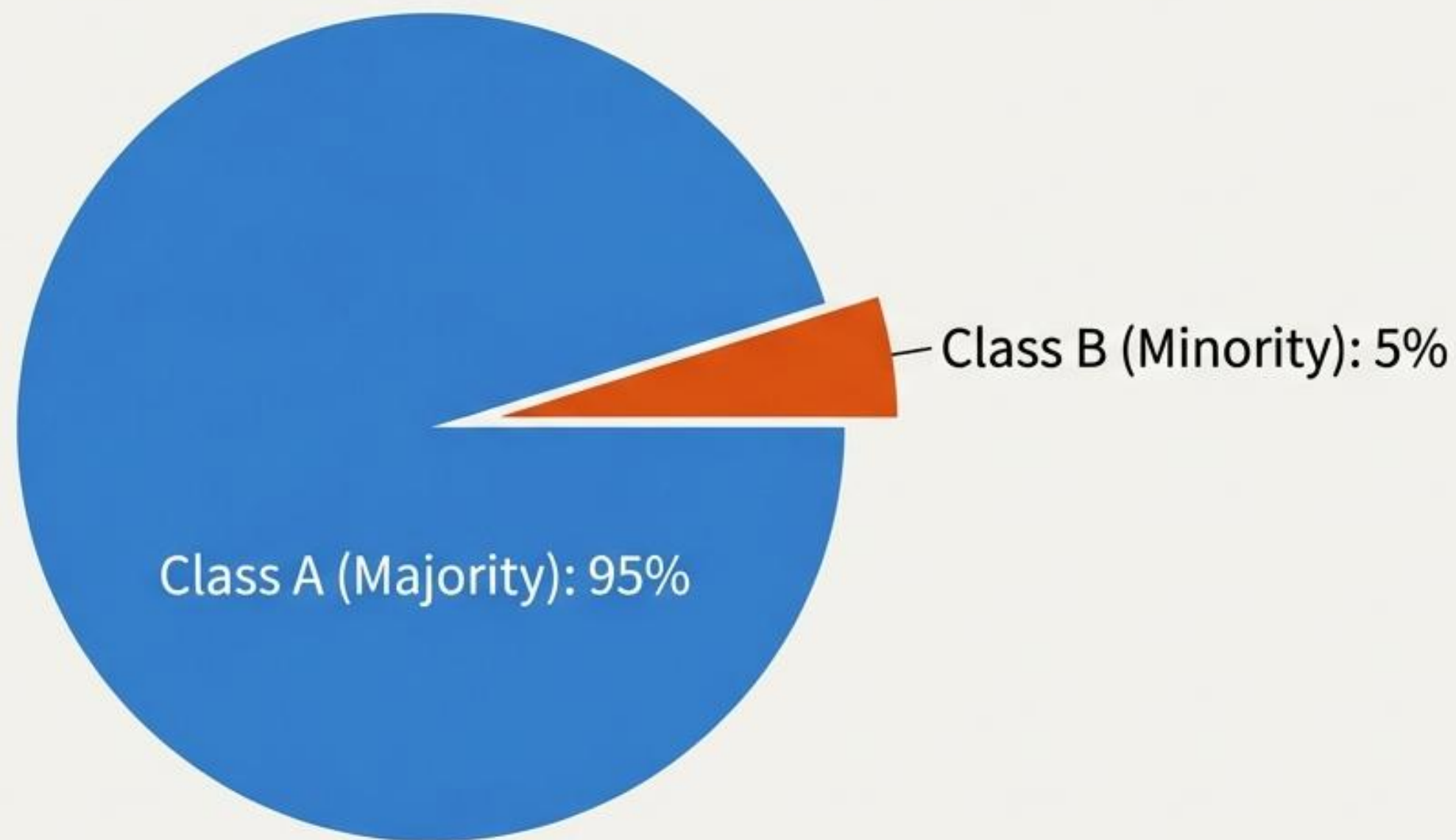
三個類別比例完全相同
(33% / 33% / 33%)。

屬於「平衡資料集」
(Balanced Dataset)。



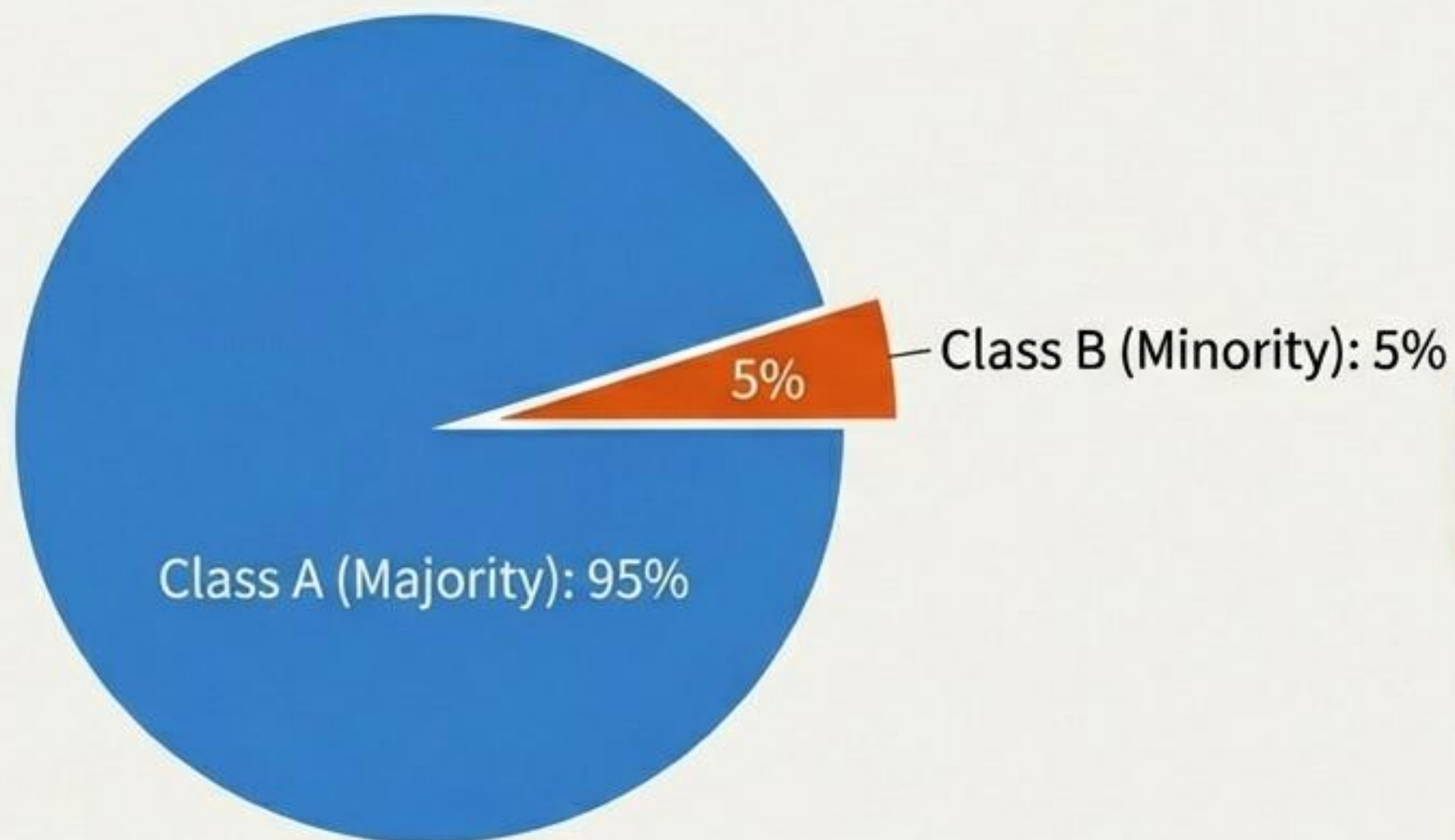
非常適合用來示範基礎分類模型。

真實世界的挑戰：類別不平衡



在真實工業場景 (如瑕疵檢測、主軸健康預診) 中，異常數據往往只佔極小比例。

準確率高就好嗎？打破 95% 的假象



Imbalanced Dataset (Real Problem)

如果模型「作弊」，盲目預測
所有資料都是 Class A...

Accuracy (準確率) = 95%

但在 **Class B (少數類別)**
的預測**成功率 = 0 !**

針對極端不平衡的資料，Accuracy 是一項不可靠的指標。我們需要全新的評估工具。

混淆矩陣 (Confusion Matrix) : 透視預測的真相

		Predicted Class (預測類別)	
		Positive	Negative
True Class (真實類別)	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

超越準確率：進階評估指標

準確度 (Accuracy)

$$\frac{(TP+TN)}{\text{Total}}$$

模型整體預測正確的比例。適合「類別平衡」的情況。

靈敏度 (Sensitivity / Recall)

$$\frac{TP}{(TP+FN)}$$

衡量模型「抓到正樣本」的能力。對漏判 (FN) 非常敏感。

精確度 (Precision)

$$\frac{TP}{(TP+FP)}$$

衡量模型「預測為正時的可靠性」。對誤判 (FP) 非常敏感。

從分類到迴歸：思維模式的根本轉折

	 分類 Classification	 迴歸 Regression
核心思維	非黑即白 (Right or Wrong)	誤差範圍 (Margin of Error)
錯誤定義	將 Setosa 誤判為 Versicolor 就是絕對的「錯」	房價 1000 萬，模型猜 1020 萬不算錯，只看「誤差有多大」
評估痛點	可依賴「準確率 (Accuracy)」	無法使用準確率，因為在連續世界中幾乎不可能 100% 命中特定數值