

111年統計學術研討會



# 常住人口估計模型之建置



行政院主計總處國勢普查處

報告人：郭庭嘉 科員

111年12月9日

# 大綱

## 01 背景說明

## 02 編算及估計方法

- 編算全國常住人口總數
- 建立常住縣市估計模型

## 03 模型評估

## 04 結語



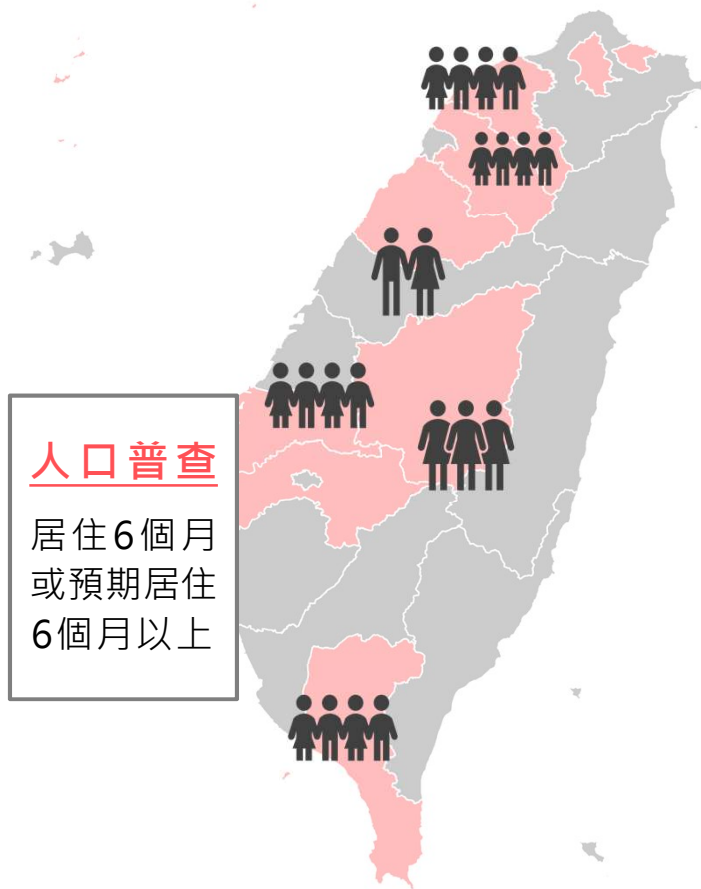
01

背景說明



# 背景說明 ▶▶

- ◆ 常住人口定義：指在某個區域範圍內實際經常居住的人口

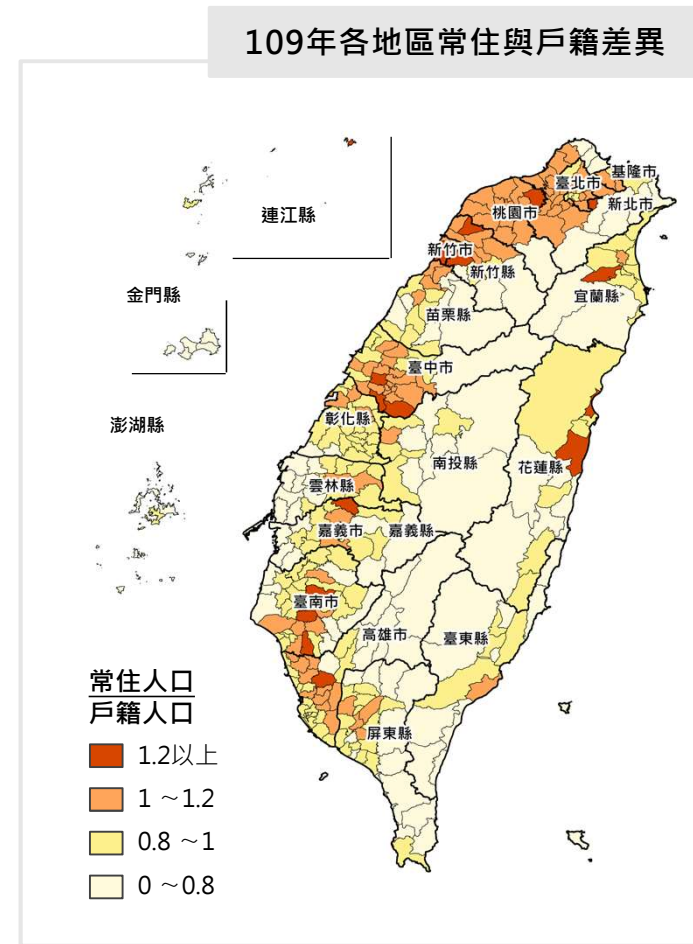
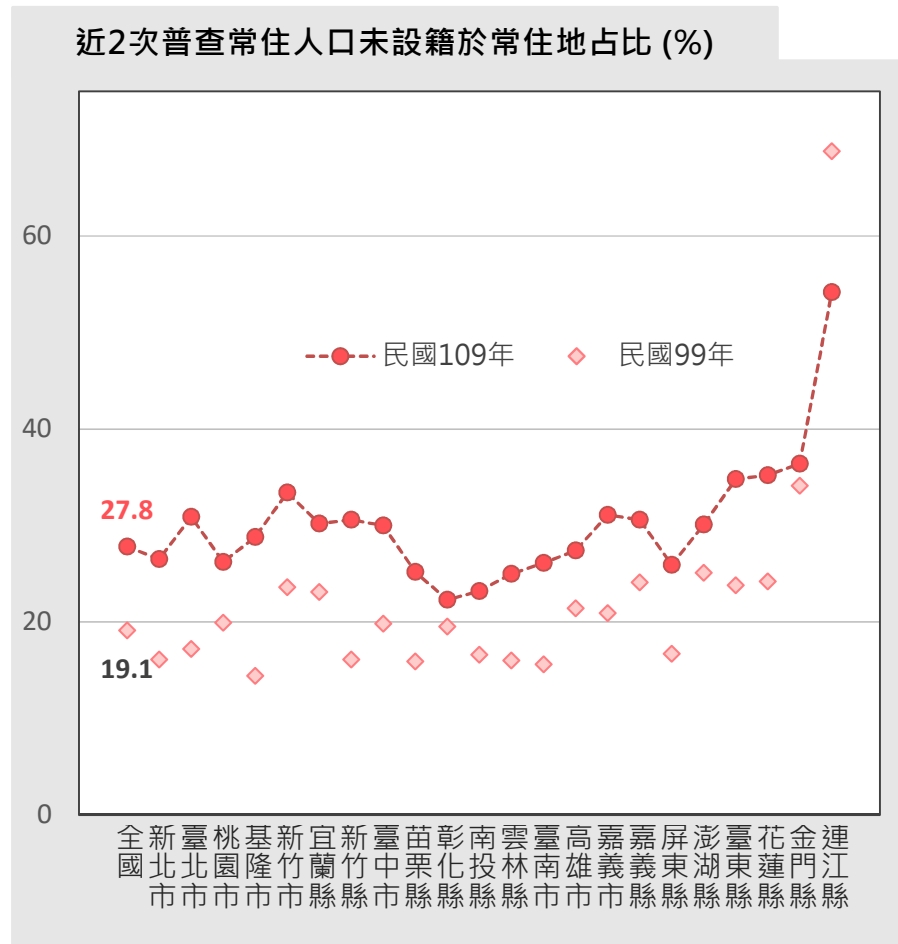


- ◆ 常住人口重要性



# 背景說明 ▶▶

## ◆ 未設戶籍於常住地占近3成，常住與戶籍差異漸大



# 背景說明 ▶▶

## ◆ 人口普查10年辦理一次



## ◆ 建置常住人口估計模型



# 02

## 編算及估計 方法

## 編算及估計方法 ▶▶

- ◆ 編算全國常住人口總數及建立常住縣市估計模型





# 編算全國常住人口總數 ▶▶

## ◆ 基準日編算法



## ◆ 基期累算法 人口變動要素合成法

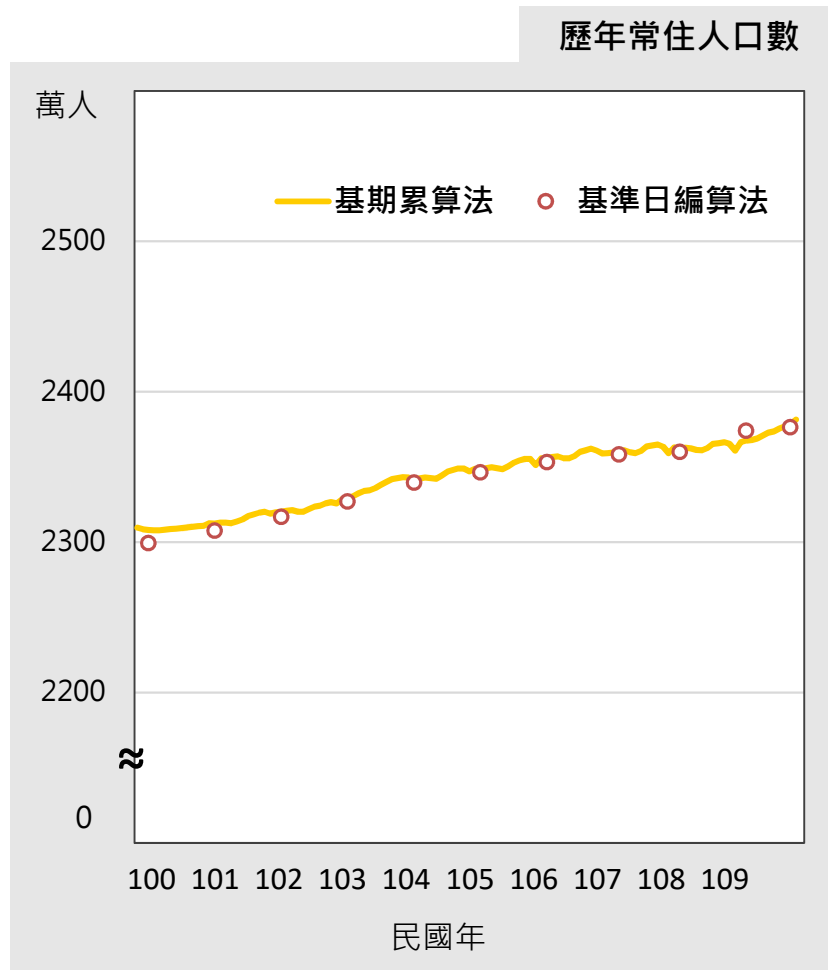
$$P_{t+1} = P_t + B_{t,t+1} - D_{t,t+1} + \text{NOM}_{t,t+1}$$

本期發生

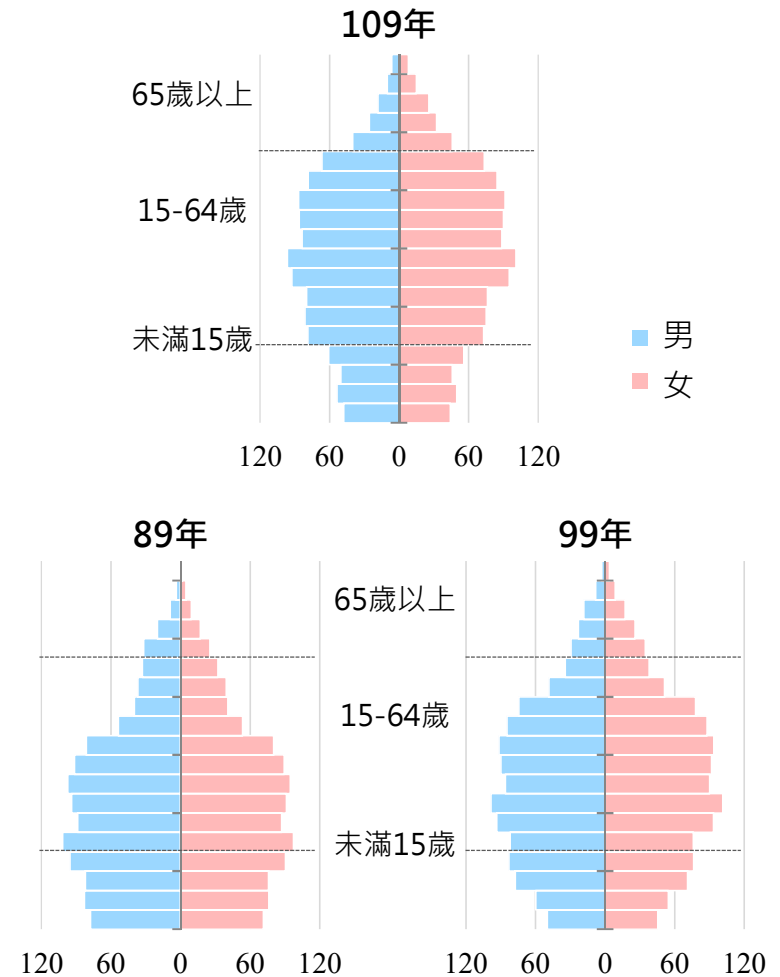


# 編算全國常住人口總數 ▶▶

## ◆ 全國常住人口總數編算結果

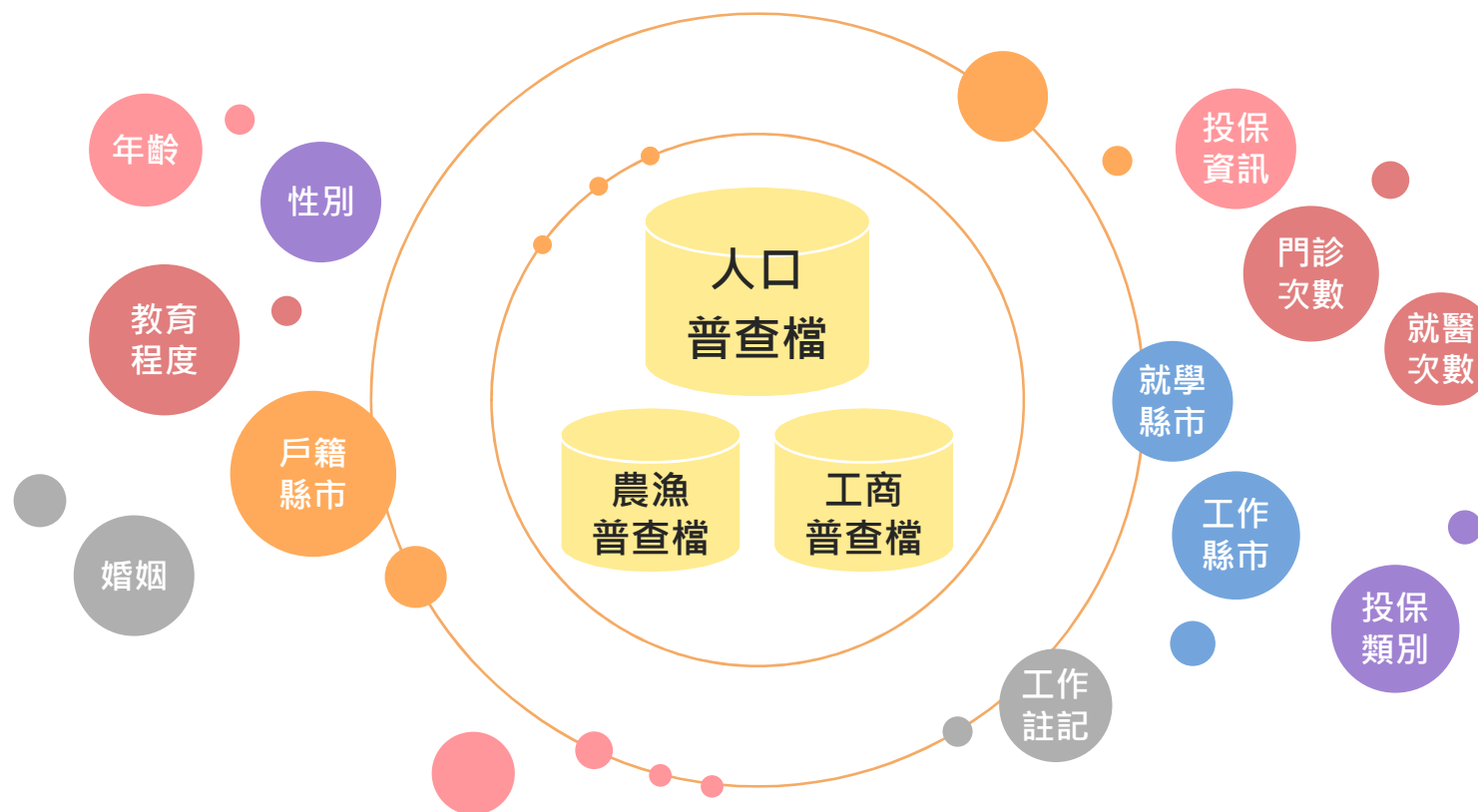


## ◆ 全國常住人口金字塔



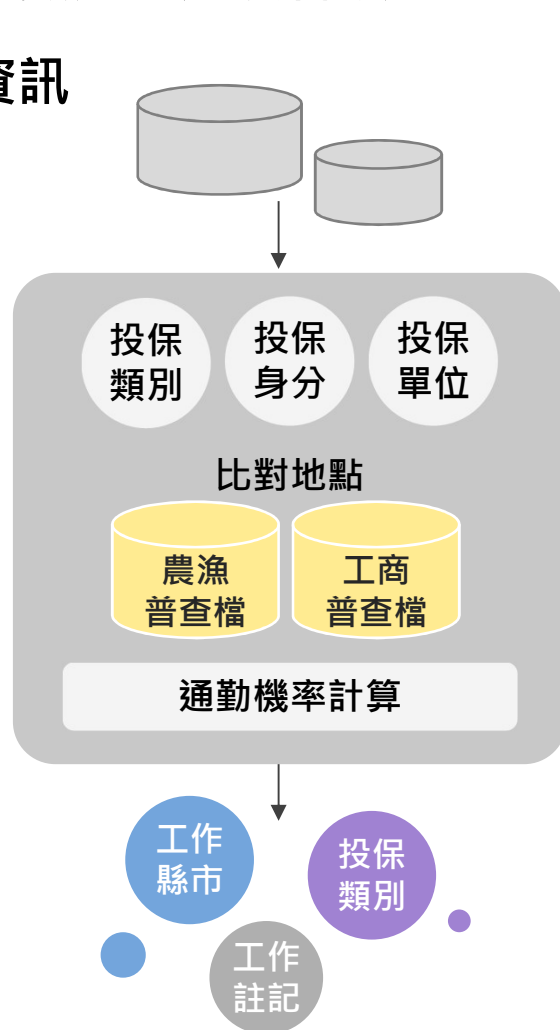
# 建立常住縣市估計模型 ▶▶

- ◆ 整合不同領域及多元屬性資料
- ◆ 資料清理、整合及資料轉換



# 建立常住縣市估計模型 ▶▶

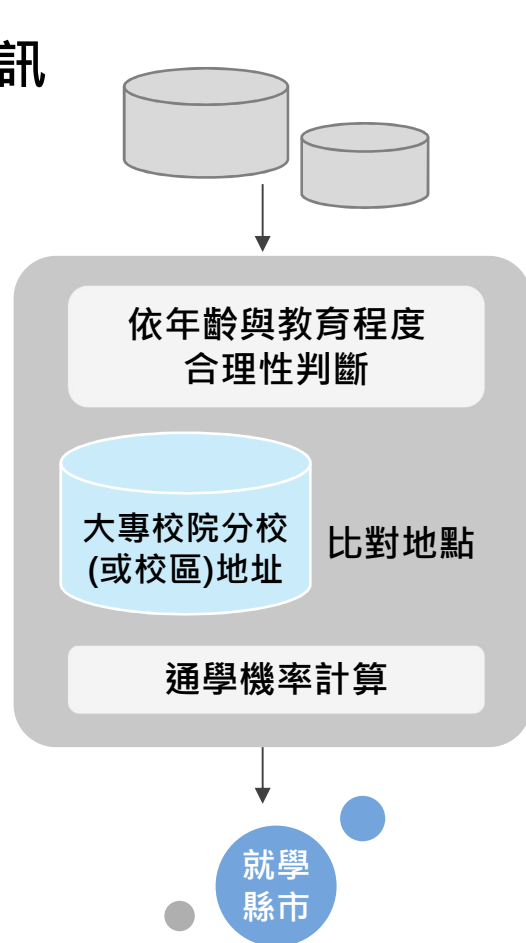
- ◆ 投入變數之篩選及轉換
- ◆ 工作資訊



# 建立常住縣市估計模型 ▶▶

◆ 投入變數之篩選及轉換

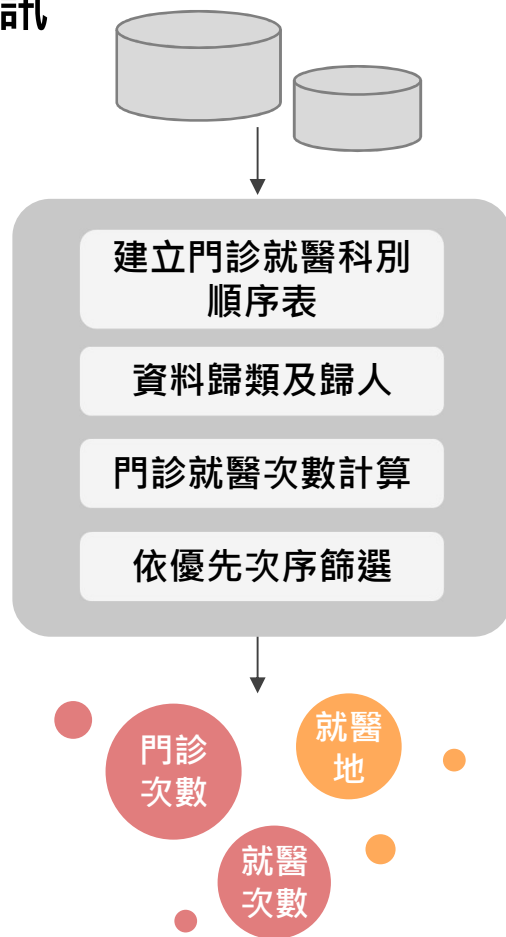
◆ 就學資訊



# 建立常住縣市估計模型 ▶▶

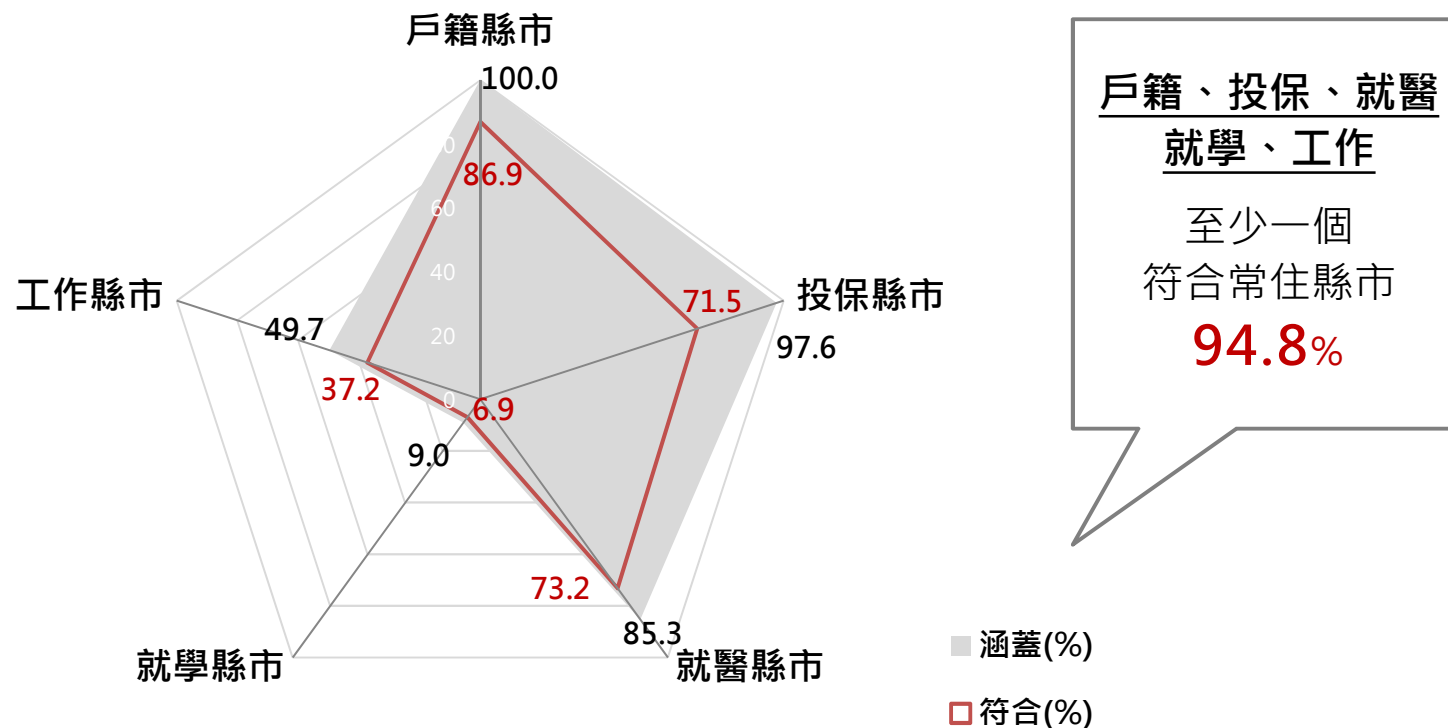
- ◆ 投入變數之篩選及轉換

- ◆ 就醫資訊



# 建立常住縣市估計模型 ▶▶

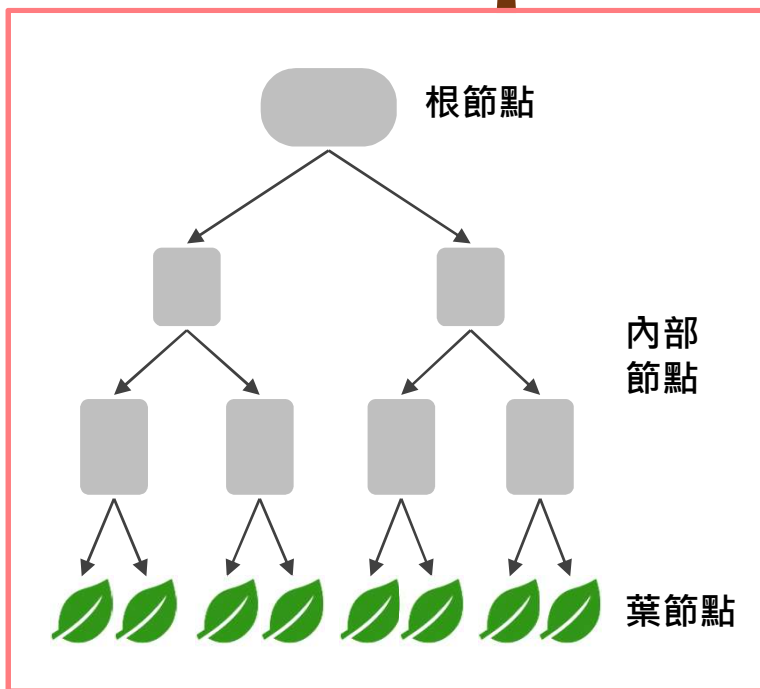
- ◆ 投入變數包含個人基本特徵及地點相關之公務登記資料
- ◆ 地點相關之公務登記資料涵蓋常住縣市比率高



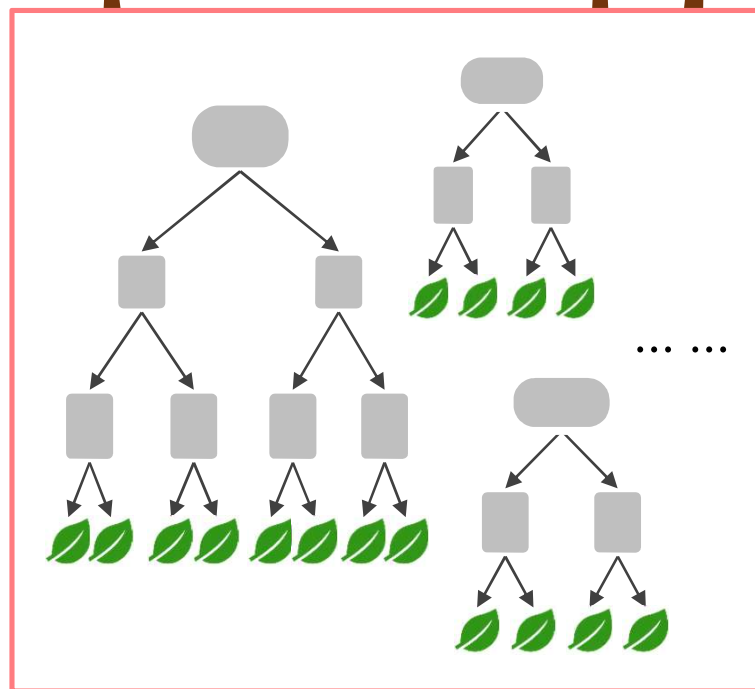
# 建立常住縣市估計模型 ▶▶

- ◆ 運用機器學習方法進行建模

決策樹



隨機森林





# 建立常住縣市估計模型 ▶▶

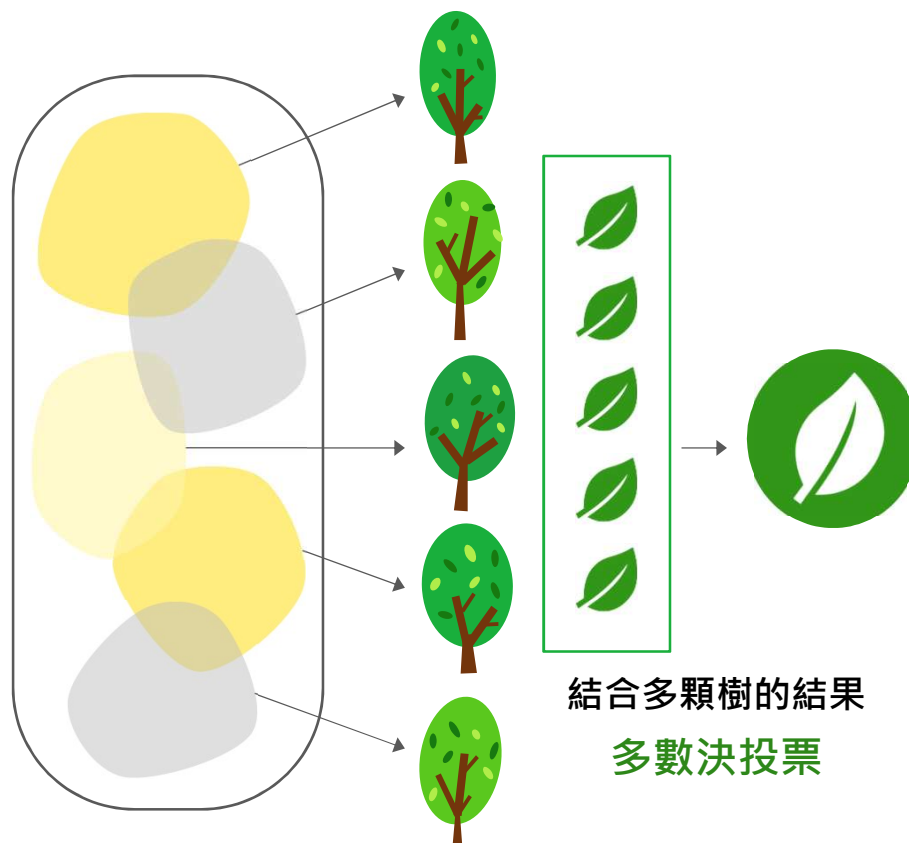
## ◆ 隨機森林優勢

採用Bagging方法的  
集成式學習  
(Ensemble learning)

採隨機抽出放回方式  
生成訓練資料集

每個訓練資料集隨機  
選擇變數

避免過度配適  
(Overfitting)

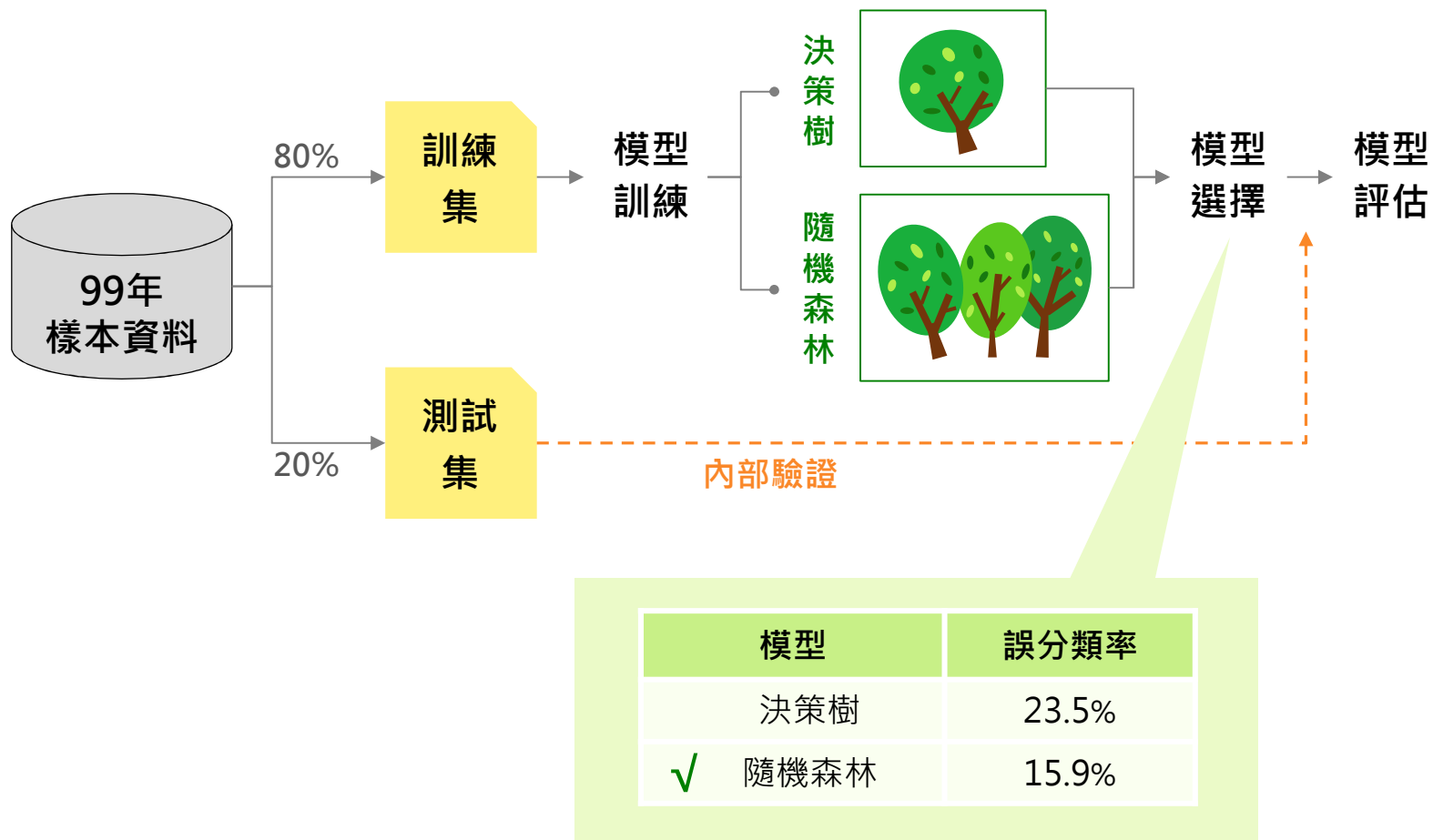


03

模型評估

# 模型評估 ▶▶

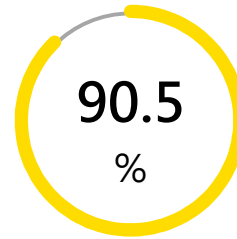
## ◆ 99年模型建置及普查資料驗證



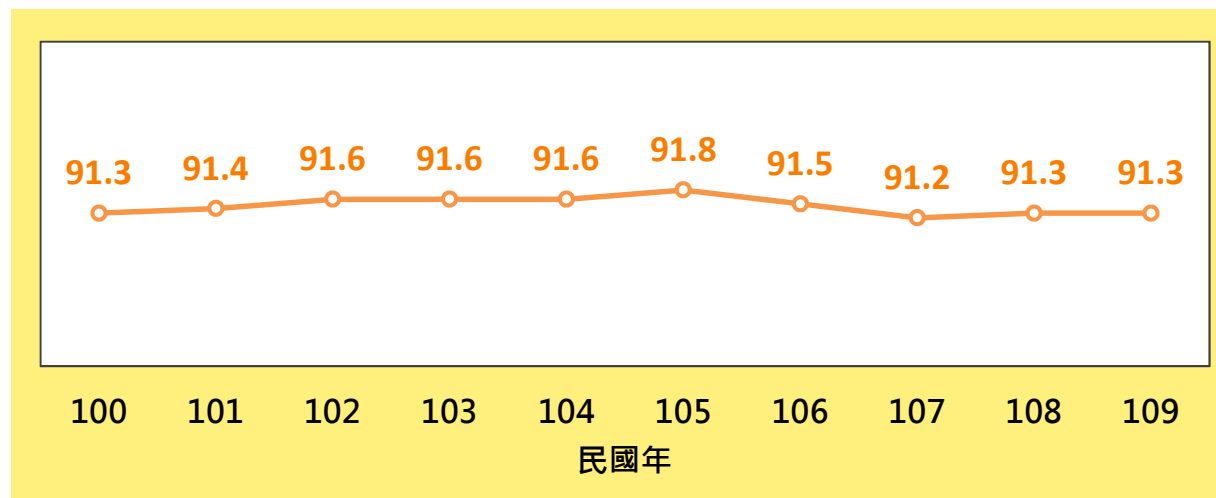
# 模型評估 ▶▶

## ◆ 99年模型驗證

內部驗證：99年測試樣本

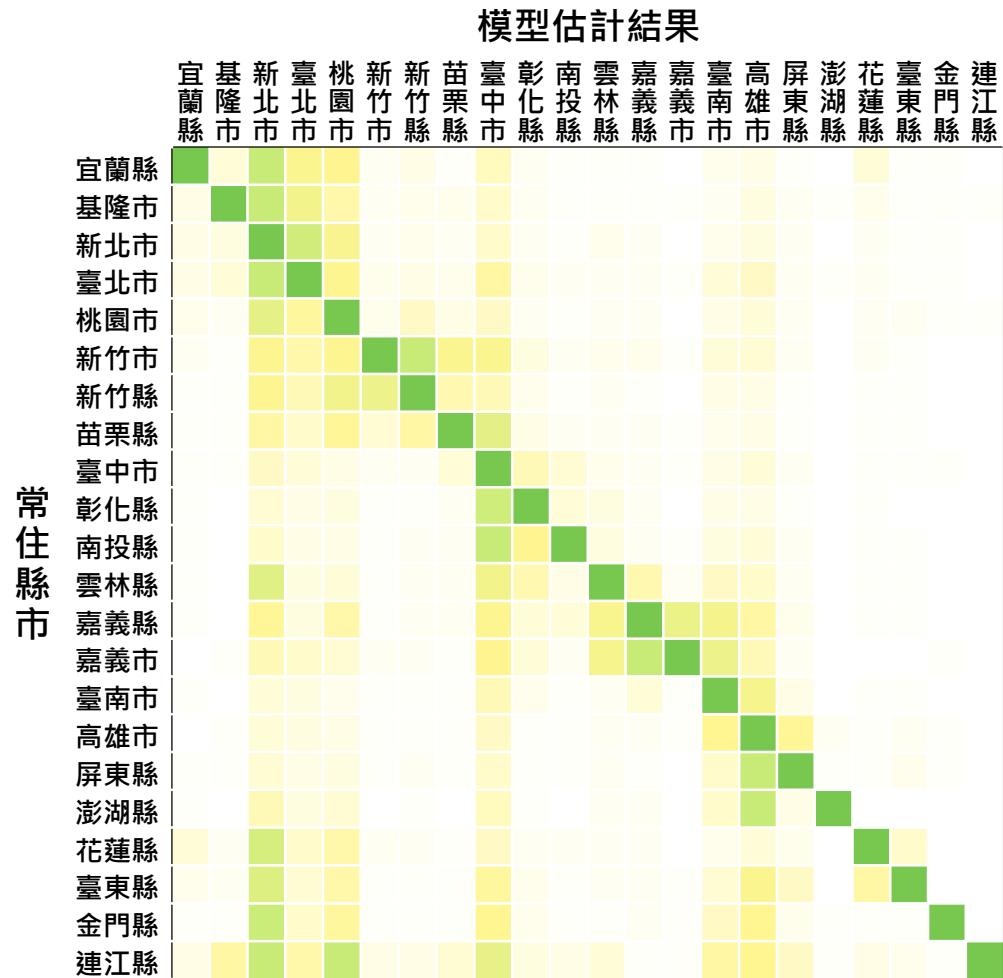
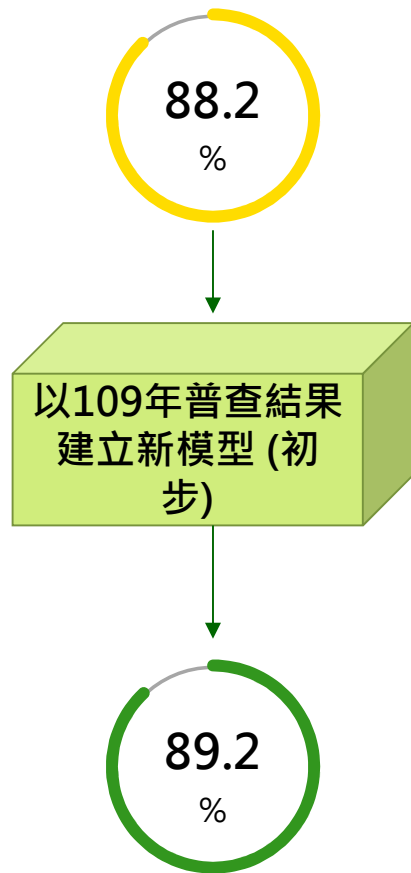


外部驗證：100年~109年人力資源調查樣本



# 模型評估 ▶▶


## ◆ 以109年普查資料驗證99年模型以及建立新模型





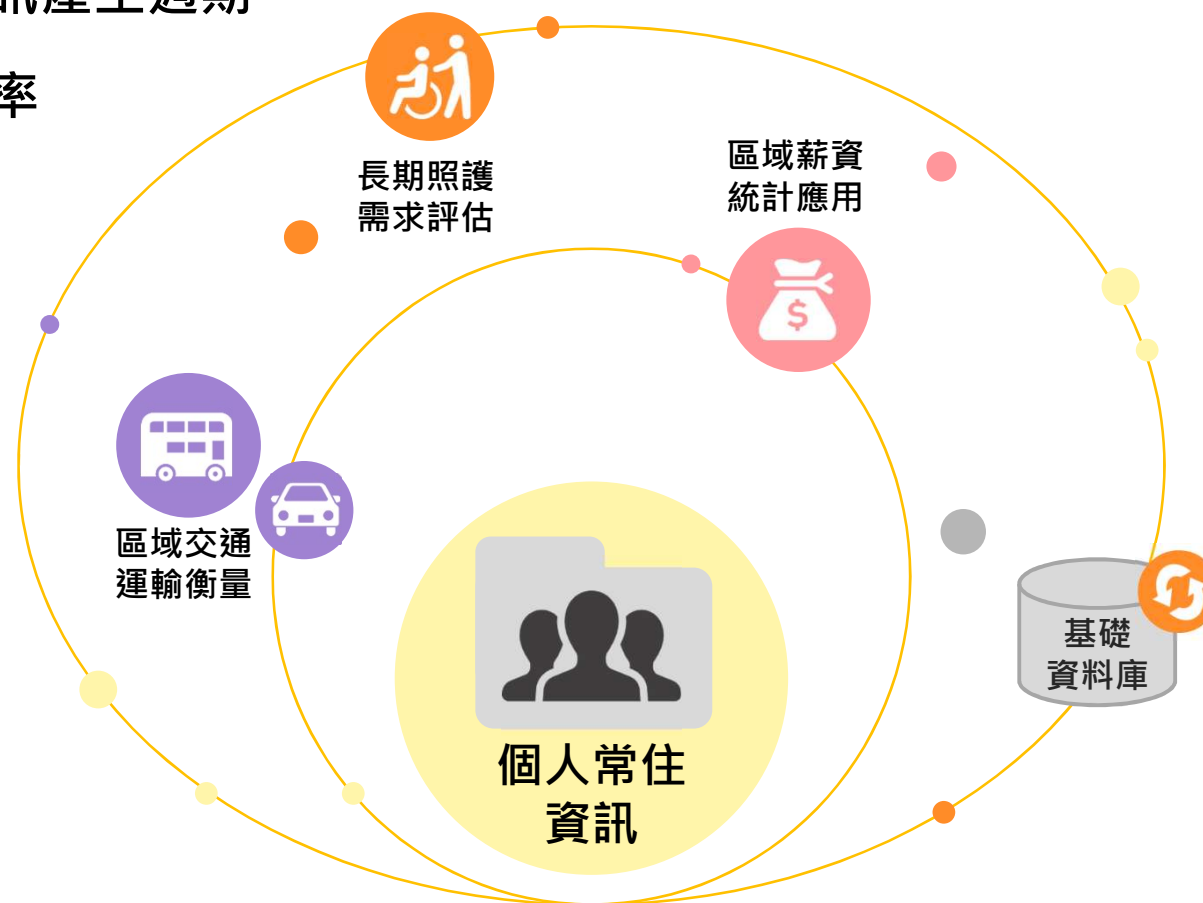
# 04

## 結語



## 結語 ▶▶

- ◆ 模型估計結果實際運用於109年普查抽樣設計
- ◆ 縮短常住人口資訊產生週期
- ◆ 提高資料產製效率
- ◆ 增廣應用範疇





簡報完畢 敬請指教

