

第 2 章 資料的種類與次級資料的蒐集

- 統計學的第一步是蒐集資料，因為資料是統計學的基礎、是統計學所要處理的對象；若沒有資料，從何統計。
- **資料的意義**：在一定的時間與空間內，人類的各項活動或大自然的現象等群體中有關人、事或物，依其個體特性來描述或記載的紀錄稱為資料(data)。
- 統計學所要處理的資料：
 - 是群體的、大量的，而非特別的個案。
 - 是可以計算或可以測量的(可數字化)。
 - 是客觀的：統計資料是經過調查、實驗、觀察、描述登錄而得到的具體事實的結果，因此統計資料是客觀的。
- 統計分析之目的在於從大量的資料中整理出有意義的資訊，所以統計數據應具有其意義，整理出數據後也應好好解釋其意義。

1

- 根據農委會的資料，2005 年家犬數量為 1,131,374 隻，較 2003 年減少 223,839 隻，降幅達 16.5%，持續下滑趨勢。(中央社，2006/1/9)

- 以上資料是他人提供的，屬次級資料；這種資料較容易取得。問題是
 - 這些統計資料是如何產生的？
 - 可信嗎？每當接觸到次級資料，需謹記這兩個問題，不要隨便相信。

資料的種類

- 依取得的方式
 - **初級資料(primary data)**：直接由研究人員或資料使用者依研究的目的去調查、觀察或實驗而獲得的資料。
 - **次級資料(secondary data)**：他人所搜集、整理分析的統計資料稱為二手資料或次級資料。

3

➢ 95 年工商及服務業普查：實地訪問 1,105,421 家企業之結果

產業別	合計	90 年以前開業 (1)	91~95 年開業 (2)	90 年底 (3)	加入率 (2)/(3)	存活率 (4)=(1)/(3)	退出率 1-(4)
製造業	147,910	116,362	31,548	140,571	22.44	82.78	17.22
傳統產業	132,848	105,006	27,842	125,768	22.14	83.49	16.51
非傳統產業	15,062	11,356	3,706	14,803	25.04	76.71	23.29
服務業	879,483	526,273	353,210	735,231	48.04	71.58	28.42
知識密集型	114,078	67,376	46,702	75,455	61.89	89.29	10.71
非知識密集型	765,405	458,897	306,508	659,776	46.46	69.55	30.45

- 製造業：非傳統產業相對於傳統產業其加入率較高，退出率亦高，顯示非傳統產業較不穩定，市場競爭性高，產業仍在發展中。
- 服務業：服務業成長速度遠高於製造業。知識密集型加入率高、退出率較低，代表知識密集型服務業持續成長；非知識密集型之加入率雖高，但其退出率亦高；欲從事服務業，知識密集型較可能存活。

2

- 例如：家犬數量的資料，由於是農委會自行蒐集，對農委會來說即為初級資料，但對中央社及其讀者(我們)來說是次級資料

- 依資料的性質
 - **質的資料(qualitative data)**：凡是不以數值來表示，僅以類別區分的資料，稱為質的資料，又稱為**類別資料(categorical data)**。
 - 例如：調查樣本個體的教育程度(國小、國中、高中、大學等)、性別、婚姻狀況等資料，每筆資料代表的僅是個體的某項性質，屬質的資料(類別資料)。
 - **數量資料(quantitative data)**：凡是可計數的資料稱為數量資料。
 - 例如：調查家犬數量、失業率、家庭收入、智商等資料，每筆資料都具有明確的數字，故屬數量資料。

4

● 依資料的發生時間

- **橫斷面資料(cross section data)**：發生於同一時點或同一期間的資料稱為橫斷面資料。
- 例如：調查台灣所有上市公司 2009 年年底的股票價格資料、調查台灣所有上市公司在 2009 年的總資產成長率，均屬橫斷面資料。
- **時間數列資料(time series data)**：發生於不同時點或不同期間的資料稱為時間數列資料。
- 例如：台積電在 2009 年的每日股票收盤價，這些有關一個個體在不同時間點的觀察值即形成時間序列資料。
- **縱橫資料(panel data)**：包含時間數列與橫斷面之資料稱為縱橫資料(追蹤資料)。
- 例如：調查台積電、聯電、中信金、中鋼、鴻海等五家公司在 2001~2009 年的總資產成長率，即形成縱橫資料。

5

- **未分組資料(ungrouped data)**：資料沒有被分類或分組的稱為未分組資料或稱粗資料(raw data)。

● 依資料的觀察值

- **間斷資料(discrete data)**：資料(觀察值)的個數是有限的，或個數是無限但可數的，稱為間斷資料。
- 例如：一個家庭中的小孩子個數(也許是 0,1,2,...個)、計程車司機一天的載客數(0,1,2,...個)。
- **連續資料(continuous data)**：資料(觀察值)的個數為無限且不可數的，稱為連續資料。
- 例如：等候公车的時間(1.1 分、10.5 分，有太多種無法清楚算出個數的可能性)，股票的報酬率(1.22%、-3.45%等，可能狀況也數不出來)。註：由於樣本個數是有限的，所以觀察值的可能狀況在樣本中絕對可以算的出來；連續資料這個觀念，是從抽樣前的角度來看無法計算出其可能狀況。

7

● 依資料的對象範圍

- **普查資料(census data)**：針對整個母體的每一元素進行全面性調查而得到的資料稱為普查資料。(普查資料是母體資料)
- 例如：2008 年的嬰兒出生數資料(根據戶口登記的統計結果)、中華民國台閩地區工商普查的報告均是普查資料。
- **抽樣資料(sampling data)**：由母體中所抽取的樣本而獲得的資料稱為抽樣資料。
- 例如：「企業最愛千里馬-2008 年大學生評價調查」用的就是抽樣資料。

● 依資料的呈現方式

- **分組資料(grouped data)**：凡是已經依照統計方法來分類或分組的資料稱分組資料。
- 例如：例如台灣的失業率調查，將失業者依教育程度區分為不識字、國小、國中、高中、高職、大學以上等。

6

- 如果資料是不可信的或是有問題的，那麼統計的結果或推論是不正確的或不可信的。因此，如何蒐集或利用可靠的資料，便成為統計學最初步且最重要的工作。由於次級資料是他人所提供，因此在使用次級資料前，必須先清楚了解

- 提供資料的單位是否公正客觀
- 資料是否符合研究目的
- 變數定義的內容與時效性
- 資料的蒐集方法為何
- 若資料已經整理過，那麼資料是如何產生的
- 資料可不可靠
- 數據是否合理
- 數據有無計算錯誤之疑慮

註：初級資料是自行蒐集，雖取得資料的成本較高，但比較容易控制上述問題。

8

次級資料的來源

- 政府機關：行政院主計處、中央銀行等單位，有許多跟總體經濟有關的數據；台灣證券交易所、證券櫃檯買賣中心、台灣期貨交易所等有金融商品的交易資料。
- 商業資料庫：台灣經濟新報、時報資訊情報贏家。包含上市公司股票交易資料、共同基金淨值及其相關資訊、上市公司財務報表、國際金融統計資料等。
- 網路：yahoo Finance (finance.yahoo.com)在世界各國的網站上有國內外金融資產的交易資料。

9

The screenshot displays the Government Statistical Portal (政府統計總覽) website. The left sidebar contains a navigation menu with categories such as '關於主計處', '政府預算', '政府會計', '預算執行及決算', '政府統計', '政府統計總覽', '主計處統計專區', '資訊服務', '統計法制與標準分類', '全國統計資料', '應用用途別系統', '各種物價空白快報表下載區', '固定資產建設改良擴充及變賣表', '統計圖說', '相關統計網站', and '資訊管理'. The main content area is titled '政府統計總覽' and features several sections:

- 【主計處統計專區】**: A grid of links for various statistics including '物價指數', '國民所得及經濟成長', '綠色國民所得', '家庭收支調查', '就業、失業統計', '薪資及生產力統計', '社會指標', '性別統計', '工商及服務業普查', '戶口及住宅普查', '農林漁牧業普查', '國富統計', '產業關聯統計', '社會發展趨勢調查', '人力運用調查', '婦女婚育與就業調查', '青少年狀況調查', '國內遷徙調查', '缺工及人力需求統計', '員工進退、勞動報酬統計', and '攤販概況統計', '農漁產品運銷實況統計'.
- 【資訊服務】**: Links for '經社指標', '縣市指標', '環境永續指數', '統計分析', '統計資料庫', '出版品及統計資訊提供', and '答客問'.
- 【統計法制與標準分類】**: Links for '政府統計制度與組織', '統計法規', '統計標準分類', '統計發展中程計畫', '統計範圍劃分方案', and '統計調查管理作業事項'.
- 【全國統計資料】**: A grid of links for '地理及環境', '人口', '勞工', '教育、科學、文化及大眾傳播', '衛生', '司法、公共秩序及安全', '社會保險及福利', '國民經濟', '農林漁牧業', '工礦業', '商業貿易', '運輸、倉儲及通信業', and '財政金融', '物價', '一般政務', '其他'.

10