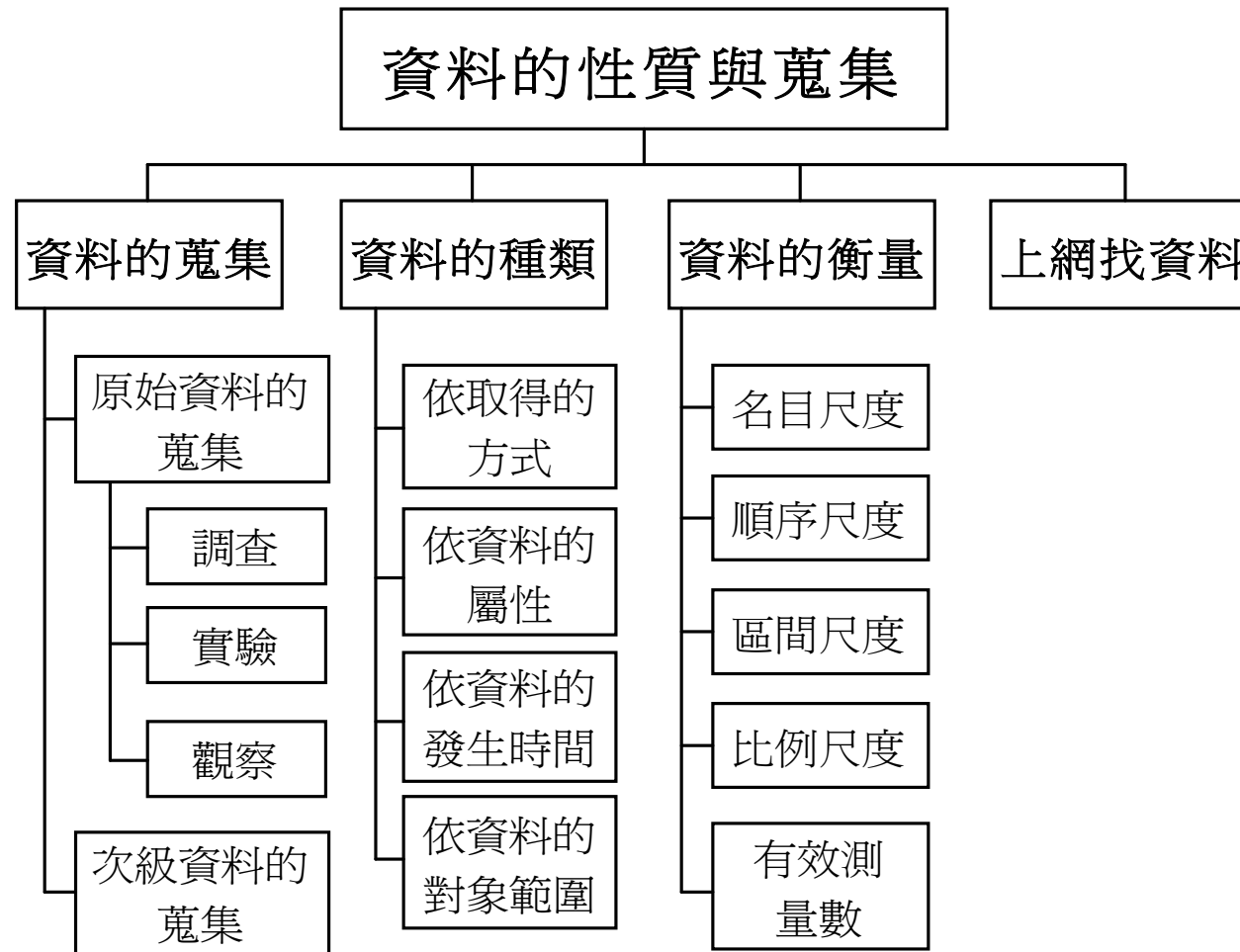


第2章 資料的性質與蒐集

學習目的

1. 了解資料的意義、性質與種類。
2. 學習不同種類資料的搜集方法，包括調查、觀察與實驗等。
3. 知悉資料的來源。
4. 了解搜集資料時的應注意事項。
5. 區分四種資料的衡量尺度及其使用。
6. 了解調查的實施及執行方式。

本章結構



資料的意義

○ 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

- ①是群體的、大量的，而非特別的個案。
- ②是可計可量的。
- ③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。

Example

- ◆ 台灣電子業廠商營業收入的規模分配為何？
- ◆ 台灣15歲以上勞動力的失業情形為何？
- ◆ 歷年臺灣颱風的次數與降雨量為何？
- ◆ 在中大宵夜街開一家網路咖啡廳,會賺錢嗎？
- ◆ 三個促銷方案的效果,何者較佳？

2.1 資料搜集的方法

○ 原始資料的搜集方法

(一) 調查 (survey)

對影響母體特性的各種因素不做控制的情況下，進行資料搜集的方法稱為調查。

(1) 普查 (census)

針對母體中每個元素進行資料之蒐集的方法稱為普查。

(2) 抽樣 (sampling)

從母體中抽取一部份的元素進行資料蒐集的方法稱為抽樣。

(二) 實驗 (experiment)

對影響母體特性的某些因素或其他因素加以控制的資料蒐集方法稱為實驗。

(三) 觀察 (observation)

研究人員或其工作人員在做研究時利用觀看、查察記錄，而不與研究對象有任何接觸晤談的資料搜集方法稱為觀察。

說明

◆ 普查

優點: 普遍性, 全面性

缺點: 規模大, 費用高, 所需人力以多

→ 通常由政府舉辦.

EX: 工商及服務業普查, 農林漁牧業普查

◆ 抽樣

範圍小, 成本較低, 較容易進行

方法: 1. 簡單隨機抽樣 (simple random sampling)

2. 分層抽樣法 (stratified sampling)

3. 部落抽樣法 (cluster sampling)

4. 系統抽樣法 (systematic sampling)

5. 其他

說明

- ◆ 調查的實施方式與執行
 1. 通訊調查
 2. 實地個別訪問
 3. 電話訪問
 4. 受查人自行填報法
 5. 網路調查
- ◆ 問卷或調查表的設計與編製
 1. 封閉式 (主要)
 2. 開放式 (輔助)

實驗：想研究心臟病患者的生理與一般狀況

表2.1 心臟病患者的抽樣資料

	A	B	C	D	E	F	G	H	I
1		年齡	性別	身高	體重	體溫	收縮壓	教育程度	是否抽煙
2	李文正	35	男	170cm	65	37	115mmHG	專科	是
3	鍾新興	26	男	180	90	36.8	123	大學	否
4	劉會妹	45	女	158	42	36.7	130	高中	否
5	周國強	55	男	168	70	37.5	156	小學	是
6	林滿虹	52	女	165	59	37.2	125	國中	是
7	陳自立	31	男	172	71	37.3	118	研究所	否
8	張瓊珠	23	女	162	50	37.9	112	大學	否
9	賴文珍	35	女	152	39	38	135	專科	是
10	侯茜茜	58	女	150	37	36.9	127	國中	否
11	黃錦章	47	男	165	66	36	122	高中	是
12	洪耀興	37	男	182	80	37.7	145	大學	否

2.1.2 次級資料的搜集

(1) 政府機關

政府機關是最重要的次級資料的來源

(2) 學術機構

期刊, 論文 (如中央圖書館的全國學術期刊網路)

(3). 企業單位

各公司行號的經營資料, 如資產負債表

(4). 個人

個別研究人員所調查或整理的資料, 數量少且不易取得

觀念與思考

◆ 拿到資料之後怎麼辦？

1. 資料是誰提供的？公正客觀嗎？
2. 調查的母體是什麼？
3. 抽樣的方法為何？樣本有多大？抽樣誤差為多少？
4. 有效樣本數是多少？回答率為何？
5. 調查的方法是什麼？電話訪問，還是郵寄問卷或實地訪談？
6. 何時做的調查？事件發生前或發生後？距離多久？
7. 問卷的問題為何？

◆ 數據合不合理

Ex: 颱風的雨量問題

圖2.1 中央氣象局台灣歷年颱風的資料

The screenshot shows the website of the Central Weather Bureau (CWB) of Taiwan. The page title is "歷年颱風基本資料表" (Historical Typhoon Basic Information Table). The table lists typhoon data for the years 1958 and 1959. The table has columns for Year, Month, Date, Typhoon Name, and other statistics. A sidebar on the left contains navigation buttons for "最新消息" (Latest News), "今日颱風" (Today's Typhoon), and "專家查詢" (Expert Query). A large green text box on the right side of the page reads "請點選左表內所列之各歷史颱風以獲取資料" (Please click on the historical typhoons listed in the table on the left to obtain data).

年	月	日期	颱風名稱	Wind Peak	Pressure
1958	7	13-16	WINNIE, 溫妮	150	930
1958	8	28-30	TSORGI	40	998
1958	9	3-4	GRACE, 葛瑞絲	165	955
1959	7	14-16	BILL, 比爾	100	960

2.2 資料的種類

2.2.1 依取得的方式

原始資料

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。

二手資料

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。

資料的種類

2.2.2 依資料的屬性

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)。

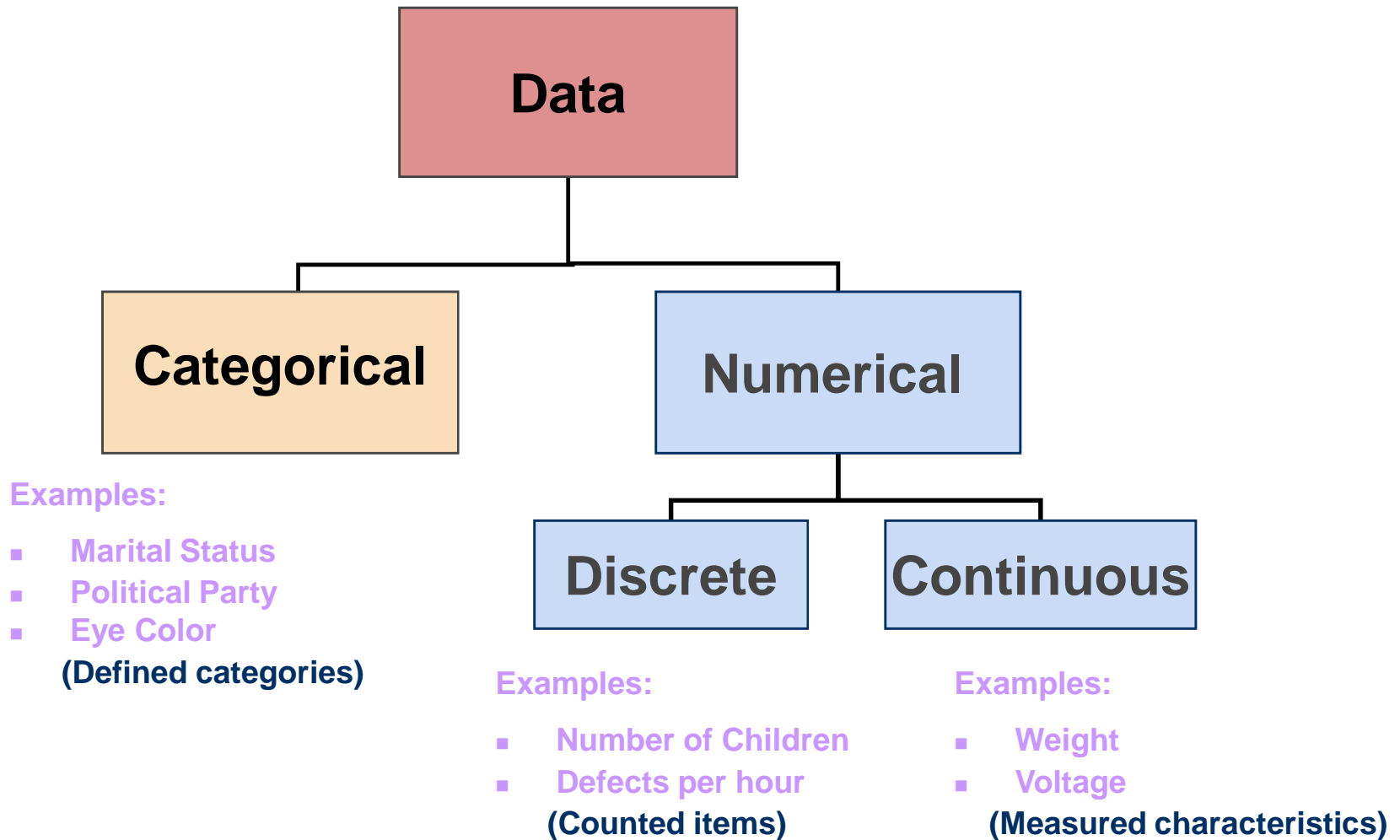
EX: 性別, 教育程度, 是否抽煙

數量資料 (quantitative data)

凡是可計數的資料稱為數量資料。

Ex: 年齡, 身高, 體重, 薪資

Types of Data



資料的種類

2.2.3 依資料的發生時間

橫斷面資料 (cross-sectional data)

發生於同一時點或同一期間的資料稱為橫斷面資料。

EX: 92年底全台灣地區商業銀行的資產資料

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。

EX: 中壢市1990~2006年的颱風資料

資料的種類

2.2.4 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。

EX: 台閩地區工商普查報告

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。

EX: 表2.1 由4000份病歷中抽取的

2.3 資料的衡量尺度

○名目尺度 (nominal scale)

衡量類別資料的稱為名目尺度又稱為類別尺度。

○順序尺度 (ordinal scale)

衡量有重要、強弱、好壞程度等級順序之資料的稱為順序尺度。

○區間尺度 (interval scale)

衡量無真正的原點的『量』的資料的稱為區間尺度。

○比例尺度 (ratio scale)

衡量有固定原點的量的資料者稱為比例尺度。

說明

◆ 名目尺度

在衡量或處理類別資料時，必須數值化才能計算比較分析
此數值只區分類別，沒有大小，順序及比例的關係

EX: 性別, 男(1), 女 (0)

◆ 順序尺度

此數值大小僅表示等級順序，數值間的差異無意義，亦不必等距

EX: 教育程度 1:小學 2:國中 3:高中 4:大學 5:研究所

◆ 區間尺度

此種資料可任意設置原點(0)，有相等或固定的衡量單位，具有大小的意義，數值之差距有意義，但其比例無意義

EX: 智商, 溫度, 時間

◆ 比例尺度

有固定原點，0代表沒有，數值有大小順序，比例值亦有意義

表2.2 心臟病患者抽樣資料的數值化

	A	B	C	D	E	F	G	H	I
1	編號	年齡	性別	身高	體重	體溫	收縮壓	教育程度	是否抽煙
2	1	35	0	170cm	65	37	115mmHG	4	1
3	2	26	0	180	90	36.8	123	5	0
4	3	45	1	158	42	36.7	130	3	0
5	4	55	0	168	70	37.5	156	1	1
6	5	52	1	165	59	37.2	125	2	1
7	6	31	0	172	71	37.3	118	6	0
8	7	23	1	162	50	37.9	112	5	0
9	8	35	1	152	39	38	135	4	1
10	9	58	1	150	37	36.9	127	2	0
11	10	47	0	165	66	36	122	3	1
12	11	37	0	182	80	37.7	145	5	0

◆ 有效測量數

當變數與所要衡量的性質有關, 或適宜代表所要衡量的性質時, 則稱此變數為有效測量數(valid measurement)

◆ 減少衡量誤差的方法

衡量誤差 = 衡量的數值 - 真值 = 偏差 + 隨機誤差

圖2.2 四種衡量尺度的比較

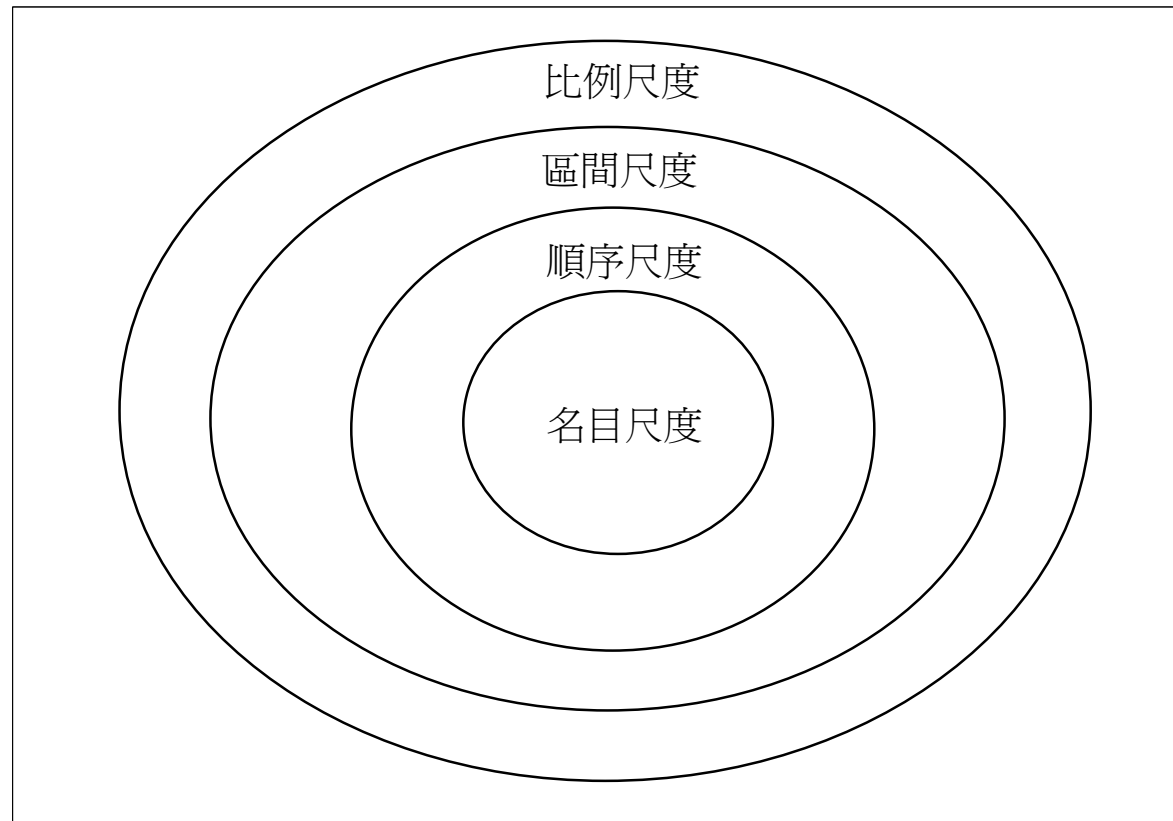


圖2.3 蕃薯藤搜尋引擎

