



統計基本名詞 與資料型態

基本名詞介紹

● 母體與樣本

- 母體(population)：欲研究之全部對象所成之集合，其範圍視研究對象而定，以本例而言，母體係指大台北地區全部行動手機使用者。
- 樣本(sample)：母體之部分集合，以本例而言，樣本係指接受調查訪問之行動手機使用者。

基本名詞介紹

- 母數 (參數) (parameter) : 用來描述母體之特徵，如母體平均值、母體標準差等。以本例而言，包含大台北地區全部行動手機使用者之平均所得、平均年齡、平均月通話費、平均購買預算等。
- 統計量 (statistics) : 用來描述樣本之特徵。例如：受訪對象之平均所得、平均年齡、平均月通話費、平均購買預算等樣本平均值。

母體與樣本之間的關係

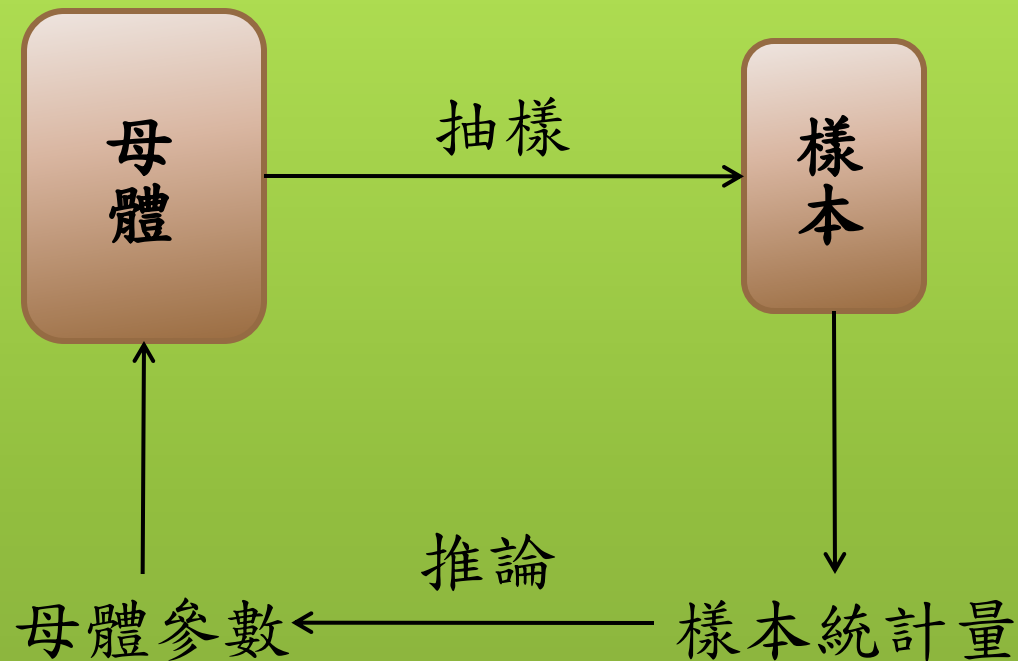
例題 1

- ◎若某研究者想要調查東門國小新生之平均身高，於是由東門國小新生入學之體檢資料得知所有新生之身高，則：
- ◎母體：所有新生之身高所成之集合。
- ◎母體參數：所有新生身高之平均數。

例題 2

- ◎若產品數量為多數，則可進行抽樣，若抽取1000件中發現有20件不良品，則可得產品之不良率為2%，則：
- ◎樣本統計量：2%。
- ◎此數值為估計產品不良率之估計量。

母體與樣本之間的關係圖



母體與樣本間的關係

基本名詞介紹

- 敘述統計與統計推論

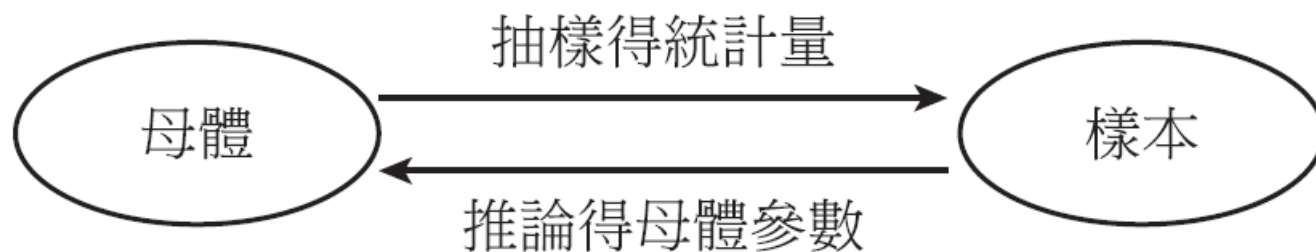
- 敘述統計 (descriptive statistics)：僅就統計資料本身特性加以描述，未將其意義擴大至更大範圍。例如：將上述手機問卷資料，整理分析受訪對象之平均所得、平均年齡、平均月通話費、平均購買預算等，並加以整理成表格與圖形，此即為敘述統計之範疇。

基本名詞介紹

- 統計推論(inferential statistics)：根據部分資料（通常為樣本）之分析結果，對更大範圍（通常為母體）之特性進行合理之推論。例如：根據上述手機問卷資料，整理分析受訪對象之平均月通話費，並進一步推論大台北地區全部行動手機使用者（母體）之平均月通話費，此即為統計推論之範疇。

基本名詞介紹

上述母體、樣本、樣本統計量、母體參數之關係如下圖。亦即針對母體抽樣而得統計量（敘述統計），再根據該統計量推論而得母體參數（統計推論）。



資料型態

- 變數與變量（值）
 - 統計資料一般係由一個以上之變數所組成，變數 (variable) 係指可加以計量或衡量之特徵，變數之值則可稱為變量（值） (value)。以手機問卷為例，性別、所得、目前使用手機品牌、通話量等，皆屬於「變數」。各變數皆可依研究目的賦予不同值，例如：性別變數值為0（代表女生）或1（代表男生）。

資料型態

- 資料衡量尺度
 - 名目尺度(nominal scale)：衡量之數字僅代表類別，並無大小或順序之意義。例如：性別變數，以0代表女生，1代表男生。名目尺度資料不可做加減乘除之運算。
 - 順序尺度(ordinal scale)：衡量之數字可代表大小或優劣順序，但數字之差距不代表其差異之程度。
 - 等距尺度(interval scale)：又稱區間尺度，衡量之數字可代表大小或優劣順序，並且數字之差距可代表其差異之程度。
 - 比率尺度(ratio scale)：當變數之衡量尺度具有等距尺度之所有性質（衡量之數字可代表大小或優劣順序，數字之差距可代表其差異之程度），並且資料間之比值具有意義。

名目尺度

- 名目尺度(Nominal Scale)又稱為類別尺度，係指用以辨識或歸類而指定的數值，可用來衡量資料的類別或型態。
- 是非題、二分題，以及複選題的選項，皆屬於名目尺度。
- 若單選題的選項為不同類別，則屬於名目尺度，如職業、血型、教育程度等問題皆屬之。
- 名目尺度的數值無大小之分，亦不能進行四則運算。例如，令男性=1，女性=0，但不能根據數值認為「男性」大於「女性」，亦不能做「男性」加減乘除「女性」的運算。

順序尺度

- 順序尺度 (Ordinal Scale) 又稱為等級尺度，係指用以排序而指定的數值，可用來衡量資料的先後、強弱、好壞、等級等關係。
- 例如，同學依成績排名被指定為第1名、第2名、第3名等，可分別編碼為1、2、3。
- 順序尺度的數值僅代表等級或順序，兩兩等級之間的距離卻不相等，故無法進行四則運算。例如，第1名與第2名的成績差距，相較於第2名與第3名的成績差距，二者並不相等。

等距尺度

- 等距尺度(Interval Scale)又稱為區間尺度，用以表示所測事物或屬性間的相對差異程度，且兩兩程度間為等距，或假設等距，如溫度。
- 具有任意原點(Arbitrary Origin)之特質，數值「0」不代表「無」或「沒有」，僅有操作上的定義。例如，溫度的攝氏0度為「水結冰的溫度」，而不是「沒有溫度」。
- 等距尺度因為數值間隔相等或固定，故能進行加減運算；但因為任意原點的性質，而不能作乘除運算。
- 李克特量表(Likert Scale)是問卷調查經常使用的問題類型，通常被假設為等距尺度。

比率尺度

- 比率尺度(Ratio Scale)也稱為比例尺度，用以表示所測事物或屬性間的實質的差異程度，如所得、銷售額、考試分數等，具有衡量單位。
- 具有絕對原點(Absolute Origin)的性質。絕對原點係指數值「0」代表「無」或「沒有」，可以進行四則運算。
- 是精確度最高的衡量尺度，亦是揭露最多資訊的詢問方式。能夠提高資料分析的準確性。
- 受訪者可能會為了避免揭露太多訊息，或是因為記憶缺乏，而不願或無法回答比率尺度的選項。因此，為提高受訪者的回答意願，有的時候寧可使用較不精確的衡量尺度。

變數種類	變數符號	衡量尺度	定義	特色	例子
分類性變數 (categorical variable)	A、B、 C....	名目 (Nominal)	用以辨識不同組別之數值	<ul style="list-style-type: none"> (1) 僅名目本身有意義，數字大小沒有意義 (2) 最低階的尺度，無法再做其他尺度轉換 (3) 僅能命名，不能排序 (4) 名目尺度誤差成本最高，衡量成本最低，總成本偏高 (5) 名目資料可以藉由百分位序升級為順序資料 	性別、職業、居住區域、身分證ID學號... 等
		順序 (Ordinal)	用以表示等級或順序之數值	<ul style="list-style-type: none"> (1) 只能指出等級或順序，但無法衡量等級間的參異程度 (2) 可降低為N尺度 (3) 僅能排序，不能加減 (4) 順序尺度，總成本較低 (5) 順序資料可以藉由李克特量表升級為區間資料 	名次、排序、百分位序... 等
分析性變數 (analytic varuable)	X、Y	區間 (Interval)	用以表示程度上差異之數值	<ul style="list-style-type: none"> (1) 具任意原點，0不代表「沒有」 (2) 可降低為O及N尺度 (3) 僅能加減，不能乘除 (4) 區間尺度，總成本較低 (5) 區間資料可以藉由分析性分組資料升級為比率資料 	溫度、態度、滿意度、同意度、重要度(李克特量表)... 等
		比率(Ratio)	用以衡量實質上差異之數值	<ul style="list-style-type: none"> (1) 具絕對原點，0代表「有」 (2) 可降低為I、O及N尺度 (3) 能加減乘除 (4) 比率尺度誤差成本最低，衡量成本最高，總成本亦偏高 	所得、銷售額、考試分數、玉米收穫量等

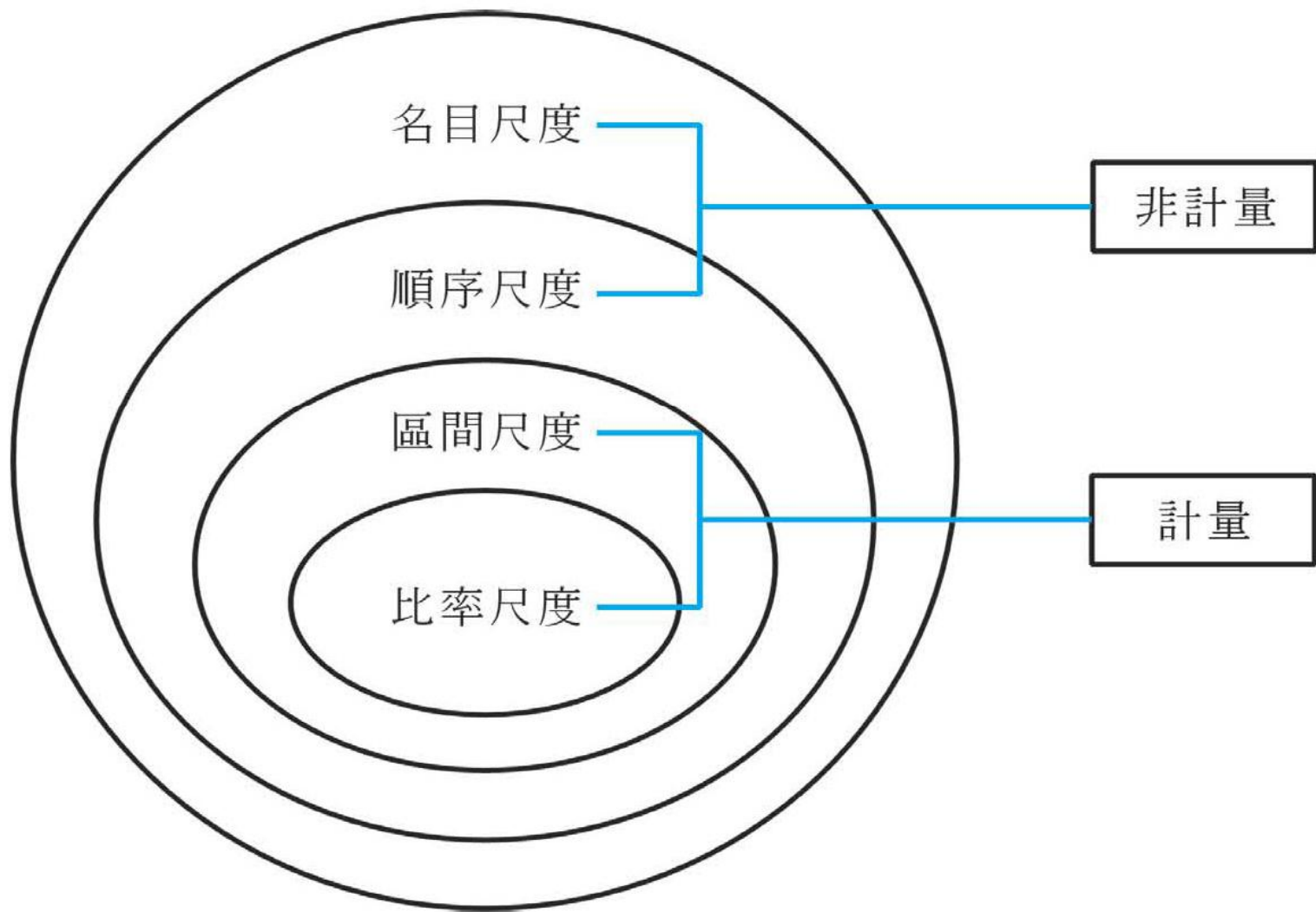


圖1-3 四個衡量尺度（名目、順序、區間及比率）的相對資訊

資料型態

- 資料 (變數) 型態：分為質化與量化資料
 - 質化資料(qualitative data)：

亦稱屬性資料，依據資料屬性或類別加以區分，又稱類別資料(categorical data)，例如：性別、居住區域、血型、膚色.....。一般而言，以名目與順序尺度加以衡量之變數皆為質化變數。
 - 量化資料(quantitative data)：

變數可用數值加以表示或計量者，例如：年齡、身高、體重、存款.....，又稱分析性資料(analytical data)。一般而言，以等距與比率尺度加以衡量之變數皆為量化變數。

資料型態

- 量化資料可進一步區分為連續與間斷資料：
 - 間斷資料(discrete data)：

具有最小計數單位之可計數資料(countable)，資料係由計數而得，其最小計數單位之間具有間隙(gap)。例如：人數、籃球得分、產品不良品數目（資料皆為整數）。
 - 連續資料(continuous data)：

其資料係由量測(measure)而得，而非計數而得，且連續資料之間具有無限多值，不存在最小計數單位或間隙。一般而言，經由物理單位量測而得之資料皆屬之，例如：身高、溫度、體重、速度、面積、時間等。 (練習)

練習

- 假設某地區之銀行，欲了解其客戶相關資料，故蒐集199份客戶資料，其蒐集資料項目（變數）與資料衡量方式（變數值）如下表。

變數名稱	變數值	衡量尺度	變數屬性
性別	1為男生；0為女生	名目尺度	質化變數
年齡	年齡值	比率尺度	量化變數
居住區域	東區為1；西區為2；中區為3	名目尺度	質化變數
信用等級	最佳為1；中等為2；不良為3	順序尺度	質化變數
職業	農漁畜牧業為1；勞工為2；軍公教為3；工商服務業為4；自由業為5；其他為6	名目尺度	質化變數
所得	年所得值（萬元）	比率尺度	量化變數
存款	存款值（萬元）	比率尺度	量化變數
貸款	貸款值（萬元）	比率尺度	量化變數
是否曾逾期繳款	有逾期繳款記錄為1；無逾期繳款記錄為0	名目尺度	質化變數

資料型態

- 資料型態與統計方法
 - 資料衡量尺度具有高低位階之分。
 - 資料衡量尺度決定變數類別（質化或量化），而變數類別又決定統計分析方法。
 - 同一變數可能因衡量方法之不同而形成不同之變數型態，進而影響統計方法與結果。

影響變數 X (自變數或獨立變數)		受影響變數 X (依變數)		統計方法
質化	性別、教育程度	量化	購買預算	變異數分析或 t 檢定
質化	手機用途	質化	手機品牌	卡方檢定
量化	年齡	量化	再購意願	相關分析或迴歸分析
量化	所得	質化	品牌	區別分析