1-1 統計概論

統計學 (Statistics) 是指對欲瞭解之問題進行資料之搜集(Collection)、整理(Organization)、表現(Collection)、分析(Analysis) 及解釋(Interpretation)，並在面臨不確定的情況下，利用樣本資料所提供之資訊，來推論(Inference)母體未知性質，從而幫助決策者做出適當決策的一種科學方法。

統計可分爲敘述統計與推論統計兩大類：

一、敘述統計學 (Descriptive Statistics)：

敘述統計將蒐集到的資料整理，以統計圖、表、文字、函數、統計量表現資料所含的訊息。

二、推論統計學 (Inferential Statistics)：

利用樣本訊息對母體未知特性進行推論(估計與假設檢定)，稱爲推論統計或歸納統計 (Inductive Statistics)。

相關名詞說明如下：

1. 母體 (Population)：
   研究對象的所有個體之集合。

2. 樣本 (Sample)：
   研究對象的部份個體之集合，即母體之部分集合。

3. 參數 (Parameter)：
   描述母體資料特性之數值，又稱母數。例如：母體平均數、母體比例、母體標準差等。

4. 統計量 (Statistics)：
   隨機樣本之函數，函數中不含未知參數，用以描述樣本資料特性之數值。例如：樣本平均數、樣本中位數、樣本標準差、樣本比例等。

5. 變數 (Variable)：對母體或樣本感興趣的特性、特徵。例如：單月營收、每股盈餘、股票代號、產業別等。

6. 資料 (Data)：針對個體的相關變數進行搜集，測量所得到的訊息。

例題 1：

In a sample of 500 students in a university, 150, or 30%, are Business majors. The 30%is an example of

【93 年政大統計系】

(A) descriptive statistics (B) statistical inference

(C) a population (D) a sample

解：(A)

例題 2：

描述母體未知特性的特徵量數稱爲 【93 年銘傳大學】

(A) 參數 (B) 統計量 (C) 變數 (D) 機率 (E) 以上皆非

解：(A)

例題 3：

某選舉候選人的競選總部，對 250,000 位有投票權的合格選民進行民調。隨機抽取 2000 位合格選民，48%表達會將票投給她。試問 48%是

(A) 樣本 (sample) (B) 參數 (parameter)

(C) 統計量 (statistic) (D) 以上皆非

【104 銘傳】

解：(C)

例題 4：

試判別下列何者爲母體參數(population parameter)?①台北市今年六月份因交通意外而死亡的人數爲 45 人。②台灣地區全體高中生近視比率的抽樣調查結果爲 82%。③消基會報導某知名廠牌製造的 3A 電池其使用壽命爲 98 小時。④某醫學研究中心發表的研究報告指出成年男子的大腦平均重量較成年女子的大腦平均重量多 1.05 盎斯⑤大同大學對今年一年級全體新生所做的調查顯示一年級學生的近視比率爲 80%

(A) ①②③④⑤　(B) ①　(C) ②③④　(D) ①⑤

【106 大同資經】

解：(D)

1.2 統計資料之搜集方法、種類與衡量尺度

1.2-1 統計資料 (Data)

對人類活動或自然現象加以觀察之記錄稱為統計資料。統計資料之搜集方法可分為調查(Survey)、觀察(Observation)、實驗(Experiment)與搜集次級資料(secondary Data)等四類。

統計資料可依資料來源，調查範圍，資料屬性發生時間與測量單位可否細分加以區分，說明如下：

一、依資料來源區分：

1. 原始資料 (Raw Data)：
   研究者利用調查，觀察或實驗等各種方法，直接由資料產生處取得之資料，又稱初級資料 (Primary Data) 或第一手資料。

2. 次級資料 (Secondary Data)：
   研究者使用經他人搜集、整理過之資料，做為資料來源，此資料稱為次級資料或二手資料。

二、依調查範圍區分：

1. 普查資料 (Census Data)：
   針對所有研究對象進行調查所取得之資料，稱為普查資料。

2. 抽樣資料 (Sampling Data)：
   針對部分研究對象進行調查所取得之資料，稱為抽樣資料，又稱抽查資料。

三、依資料屬性區分：

1. 定性資料 (Qualitative Data)：
   以性質、名稱或類別區分之資料，又稱性質資料、類別資料(Categorical data)或屬質資料。例如：性別、教育程度、滿意度、血型等。定性資料以名目

或順序尺度衡量。

2. 定量資料 (Quantitative Data)：
以計量區分之資料，又稱爲數量資料(numerical data)或屬量資料。例如：身高、體重、價格、年齡、分數等。定量資料以區間或比率尺度衡量。

四、依發生時間區分：

1. 靜態資料 (Static Data)：
發生於同一時期之資料，稱爲靜態資料或橫斷面資料 (Cross-Section Data)。例如：2003 年亞洲各國人民之平均年所得。

2. 動態資料 (Dynamic Data)：
發生於不同時期之資料，稱爲動態資料或時間數列資料 (Time-Series Data)。

五、依測量單位可否無限細分區分：

1. 離散資料 (Discrete Data)
具最小計數單位，資料爲可數的，又稱不連續資料(Discontinuous Data)。例：人數、次數、個數。

2. 連續資料 (Continuous Data)
資料爲不可數的，兩個數值間可插入無限多個數值。例如：時間、重量等。

例題 5：

友愛有線電視公司爲預測晚間新聞節目的收視率，在收視戶中加裝收視率監控器，這種資料蒐集的方法是屬：

(A) 派員調查 (B) 郵寄問卷調查 (C) 實驗性資料 (D) 觀察性資料

【104 淡江管科】

解：(D)

7.1 估計方法

尋找優良估計式，常用的方法有四種：動差法、最大概似法、最小平方法與貝氏估計法。本章僅介紹動差法與最大概似法，最小平方法於第十章介紹，貝氏估計法本書不予探討。

7.1-1 動差法 (Method of Moments)

r 階樣本原動差 $m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$，$r = 1, 2, 3.....$

r 階母體原動差 $\mu'_r = E(X^r)$，$r = 1, 2, 3, ...$

利用 $1 \sim k$ 階樣本原動差等於母體原動差的方程式，聯立解出 k 個參數，即為參數之動差估計式 (Method of Moment Estimator；簡稱 MME)。

一、一個參數之估計：

利用 $E(X) = \overline{X}$，解出 θ，該 θ 值即為 M.M.E。

二、二個參數之估計：

$$\begin{cases} E(X) = \overline{X} \\ E(X^2) = \dfrac{\sum X^2}{n} \end{cases}$$，聯立求解，解出 θ_1 與 θ_2，即為 θ_1 與 θ_2 M.M.E。

三、動差估計式之性質：

1. 動差估計式必為一致估計式。

2. 動差估計式不一定為不偏估計式。

7.1-2 最大概似法 (Maximum Likelihood Method)

概似函數 (Likelihood Function) 表示在參數值 θ 下，一組樣本實現值發生的機率。若 $(x_1,x_2,....,x_n)$ 爲母體 $f(x;\theta)$ 之一組樣本實現值，則概似函數 $L(\theta)$ 定義如下：

$$L(\theta)=f(x_1,x_2,....,x_n;\theta)=\prod_{i=1}^{n} f(x_i;\theta)$$

自參數空間中找出一數值作爲 θ 之估計值，使抽出的樣本實現值 $(x_1,x_2,......,x_n)$ 出現的可能性最大，該數值即爲 θ 之 MLE (Maximum Likelihood Estimator)。

$$\underset{\theta\in\Omega}{Max}\ L(\theta)=L(\hat{\theta})\ ，$$

一、參數爲連續型式：(L(θ) 可對 θ 微分)

1. $L(\theta)$ 爲遞增或遞減函數：

由參數空間中找出使 $L(\theta)$ 極大的數值，即爲 MLE

二、參數爲離散形式

1. $L(\theta)$ 爲遞增或遞減函數：

由參數空間中找出可使 $L(\theta)$ 極大之數值，即爲 θ 之 MLE。

2. $L(\theta)$ 非遞增或遞減函數：

令 $\hat{\theta}$ 爲 θ 之 MLE

$$\begin{cases} \dfrac{L(\hat{\theta})}{L(\hat{\theta}+1)}\geq 1 \\ \dfrac{L(\hat{\theta})}{L(\hat{\theta}-1)}\geq 1 \end{cases}$$

找出 $\hat{\theta}$ 之範圍內的數值分別代入 $L(\theta)$，可使 $L(\theta)$ 極大者，即爲 M.L.E。

最大概似估計式之性質如下：

1. M.L.E 爲唯一解時，必爲一致估計式。

2. M.L.E 爲漸近不偏估計式，但不一定爲不偏估計式。

3. 若參數 θ 之充分統計量存在，則 M.L.E 必爲充分統計量。

4. M.L.E 的不變性 (Invariance Property)：若 $\hat{\theta}$ 爲 θ 之 M.L.E，$g(\theta)$ 爲 θ 之函數，則 $g(\hat{\theta})$ 必爲 $g(\theta)$ 之 M.L.E。

例：(1) $\hat{P}$ 爲 P 之 MLE，則 $\sqrt{\hat{P}(1-\hat{P})}$ 爲 $\sqrt{P(1-P)}$ 之 MLE

(2) $\hat{\theta}$ 爲 θ 之 MLE，則 $\hat{\theta}^2$ 爲 θ^2 之 MLE

例題 1：

Consider a random sample $X_1,\ldots,X_n$ from an exponential distribution with probability density function (pdf)

$$f(x)=\frac{1}{\sigma}\exp\left(-\frac{x}{\sigma}\right),\quad x>0,\sigma>0,$$

(1) Find the maximum likelihood estimator of σ.

(2) Find the unbiased estimator of σ and verify it is unbiased.

【104 淡江統計】

解：(1) $L(\sigma)=\prod_{i=1}^{n} f(x_i)=\prod_{i=1}^{n}\left[\frac{1}{\sigma}\exp(-\frac{x_i}{\sigma})\right]=\sigma^{-n}\cdot e^{-\frac{\sum_{i=1}^{n}x_i}{\sigma}}$

$$\ln L(\sigma)=-n\ln\sigma-\frac{\sum_{i=1}^{n}x_i}{\sigma}$$

令 $\frac{\partial}{\partial\sigma}\ln L(\sigma)=0$

$$\Rightarrow \frac{-n}{\sigma}+\frac{\sum_{i=1}^{n}x_i}{\sigma^2}=0$$

$$\Rightarrow -n\sigma+\sum_{i=1}^{n}x_i=0$$

$$\Rightarrow \hat{\sigma} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$$

$$\left.\frac{\partial^2}{\partial \sigma^2} \ln L(\sigma)\right|_{\sigma=\overline{X}} = \left.\frac{n}{\sigma^2} - \frac{2\sum_{i=1}^{n} x_i}{\sigma^3}\right|_{\sigma=\overline{X}} < 0$$

故 σ 的 MLE 爲 $\overline{X}$

(2) $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\sigma = \sigma$

故 $\overline{X}$ 爲 σ 的不偏估計式

例題 2：

隨機變數 x 之機率分配爲: $f(x) = (a+1)x^{\alpha}, 0 < x < 1$ 。今由該母體中抽取出一組次數爲 n 之隨機樣本 $x_1, x_2, \ldots, x_n$,

(1) 請寫出 log likelihood function

(2) 試以最大概推定法求母數 a 之推定量。

【93 中原國貿】

解：(1) $L(\alpha) = \prod_{i=1}^{n} f(x_i) = (\alpha+1)\eta \left[\prod_{i=1}^{n} X_i\right]^{\alpha}$

$$\ln \mathrm{L}(\alpha) = \ln\left[(\alpha+1)\eta \cdot \left(\prod_{i=1}^{n} X_i\right)^{\alpha}\right] = n|n(\alpha+1) + \alpha\sum_{i=1}^{n}(l_n X_i)$$

(2) 令 $\frac{\partial |\ln L(\alpha)}{\partial \alpha} = 0$

$$\Rightarrow \frac{n}{\alpha+1} + \sum_{i=1}^{n} (\ln X_i) = 0$$

$$\Rightarrow \frac{-n}{\sum_{i=1}^{n} (\ln X_i)} = \alpha + 1$$

$$\Rightarrow \alpha = \frac{-n}{\sum_{i=1}^{n} (\ln X_i)} - 1$$