CH 9

抽樣分配

Modern Engineering Statistics

工 程 統 計

(1)由已知的樣本特徵來推測(論)(Inference)未知母體真實的特徵，則稱為**推論統計(Inferential Statistics)**。

(2)在推論母體特徵(平均值μ與標準差σ)前，須自該母體抽出樣本，此過程的方法稱為**抽樣(Sampling)**。

(3)抽樣的主要目標是在有限的時間與經費下，取得具代表性的樣本。樣本特徵即所謂的**統計量(Statistics)**。

(4)統計量乃隨機抽樣而異，因此具有不確定性。即為一隨機變數且有其特定的機率分配，稱為**抽樣分配(Sampling Distribution)**。

(5)由已知樣本統計量之平均數$\overline{X}$與標準差s，可據以推論未知母體特徵值之平均數μ與標準差σ。

9-1 平均數的抽樣分配

讓我們用表9-1來說明樣本分配的觀念。

成對學生可能的組合數為$C_2^5=10$，即每一個組合其發生機率各為0.1。

表 9-1　學生考試成績結果表

姓名	考試成績	成績點數
Dan	B	3
Ed	C	2
Fran	B	3
Gil	A	4
Ida	C	2
$\mu=2.8$	$\sigma=\sqrt{0.56}=0.7483$	

表 9-2　抽出後不還原之樣本平均值的抽樣分配

可能的平均值 $\bar{x}$	可能的學生組合	$\Pr[\overline{X}=\bar{x}]$
2.0	(Ed, Ida)	0.1
2.5	(Dan, Ed) (Dan, Ida) (Ed, Fran) (Fran, Ida)	0.4
3.0	(Dan, Fran) (Ed, Gil) (Gil, Ida)	0.3
3.5	(Dan, Gil) (Fran, Gil)	0.2

◎期望值與變異數

◎與任一隨機變數相同的，隨機變數 $\overline{X}$ 亦有其期望值與變異數。$\overline{X}$ 之期望值

$$E(\overline{X}) = \sum_{所有\ \overline{x}} \overline{x}\,\Pr[\overline{X} = \overline{x}] = 2.0(0.1) + 2.5(0.4) + 3.0(0.3) + 3.5(0.2) = 2.8$$

$$E(\overline{X}) = \mu$$

◎樣本平均數 $\overline{X}$ 的變異數可由下式求得

$$Var(\overline{X}) = E(\overline{X}^2) - E(\overline{X})^2$$

$$E(\overline{X}^2) = \sum_{all\ \overline{x}} \overline{x}^2 \Pr(\overline{X} = \overline{x})$$

$$= (2.0)^2(0.1) + (2.5)^2(0.4) + (3.0)^2(0.3) + (3.5)^2(0.2) = 8.05$$

故變異數為

$$Var(\overline{X}) = 8.05 - (2.8)^2 = 0.21$$

平均數的標準差為

$$SD(\overline{X}) = \sqrt{0.21} = 0.458$$

◎樣本平均值的標準誤

(1) 樣本統計量為隨機變數，而樣本統計量隨樣本分配型態而異，故通常將抽樣分配的標準差稱為此統計量的**標準誤**(Standard Error)。

(2) $\overline{X}$ 的標準誤可用以測度 $\overline{X}$ 與 μ 間的差異。

(3) 此變異的產生是由於抽樣本身所造成的(包括樣本大小、抽樣分配型態及母體分配型態)。

(4) $\overline{X}$ 的標準誤可以下式求得

$$\sigma_{\overline{X}} = SD(\overline{X})$$

(5) $\sigma_{\overline{X}}$ 有兩種不同的計算：

5.1 抽樣採**還原(With Replacement)**或稱放回抽樣，稱為獨立的觀測。

5.2 **不還原(Without Replacement)**或稱不放回抽樣，則為非獨立(相依)的觀測。

5.3 對母體大小而言，可分為：

(1)無限母體(當採還原抽樣，觀測值視為獨立)。

(2)有限母體(當採不還原抽樣，觀測值為相依)。

(6)隨機樣本 X 是自有限母體抽樣或不還原抽樣，

則 $\overline{X}$ 抽樣分配的標準誤：

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

(7)自無限母體抽樣(或還原抽樣)，此時 $\overline{X}$ 之標準誤為

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

9-2 母體為常態分配時 $\overline{X}$ 的抽樣分配

◎ $\overline{X}$ 的常態分配

當自 (μ,σ) 常態分配之母體中抽樣，則 $\overline{X}$ 的抽樣分配亦為常態分配，此時 $\overline{X}$ 的平均數等於母體的 μ。如果母體 σ 未知，則以樣本 s 取代，亦不影響其性質，此時 $\overline{X}$ 之標準誤為

$$\sigma_{\overline{X}} = \sigma/\sqrt{n}$$

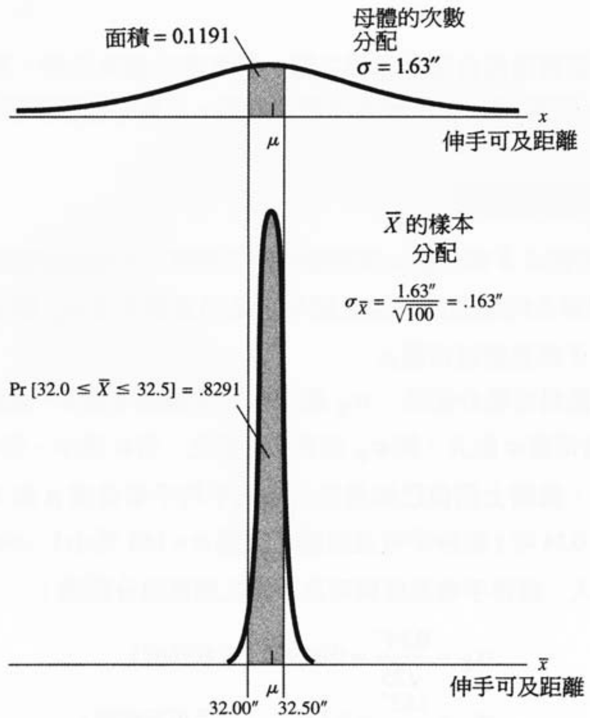

圖 9-1 母體次數分配曲線與樣本平均值的抽樣分配

◎標準誤的特性

(1) $\sigma_{\overline{X}}$ 是樣本 $\overline{X}$ 與母體 μ 間離散程度之測度，如果離散程度減少，此乃意謂著當 $\sigma_{\overline{X}}$ 減少時，任何單一樣本 $\overline{X}$ 將更接近母體 μ。

(2)當母體為常態分配時，$\sigma_{\overline{X}}$ 值取決於 σ 與樣本 n 數。

由於 $\sigma_{\overline{X}} = \sigma/\sqrt{n}$，故當母體 σ 愈大，則 $\sigma_{\overline{X}}$ 亦愈大；反之，若 σ 愈小，則 $\sigma_{\overline{X}}$ 愈小。

(3)第二個會影響 $\sigma_{\overline{X}}$ 值者為樣本大小。

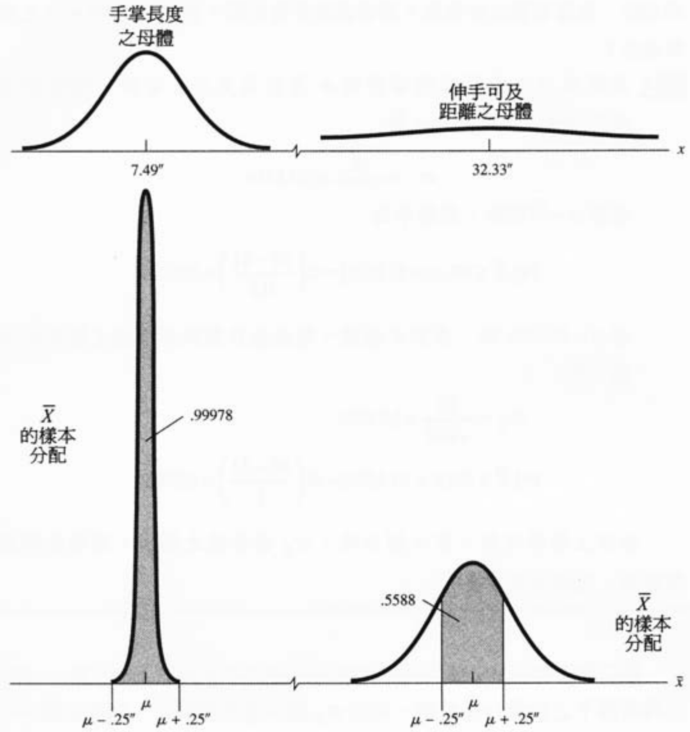

圖 9-2　成年男性手掌長度與可及距離的常態曲線

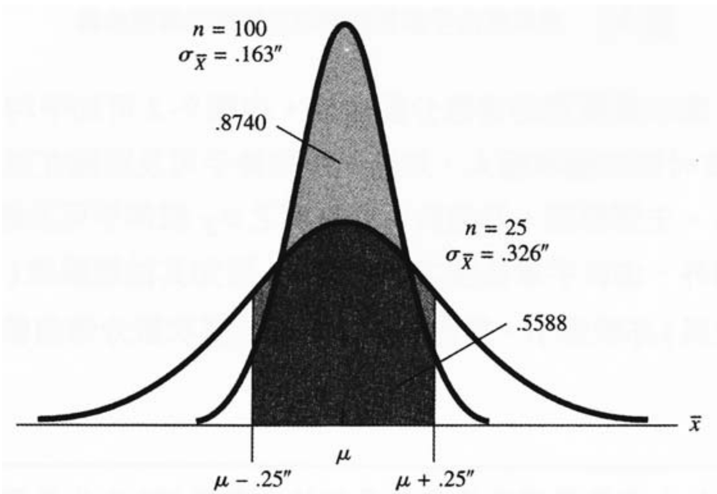

圖 9-3　平均伸手可及距離 $\bar{X}$ 的樣本分配（$n = 25$ 及 $n = 100$）

9-3 非常態母體的抽樣分配

◎若當母體並非為常態分配，只要樣本數夠大（通常指n>30），依據**中央極限定理（Central Limit Theorem）**，$\overline{X}$ 的抽樣分配會近似於常態分配。

◎中央極限定理：

所謂中央極限定理係指從一母體中隨機抽樣，不限母體的分配型態為何，只要樣本數夠大（一般為n>30），則不論母體的機率分配型態，樣本平均數 $\overline{X}$ 的抽樣分配近似於常態分配。

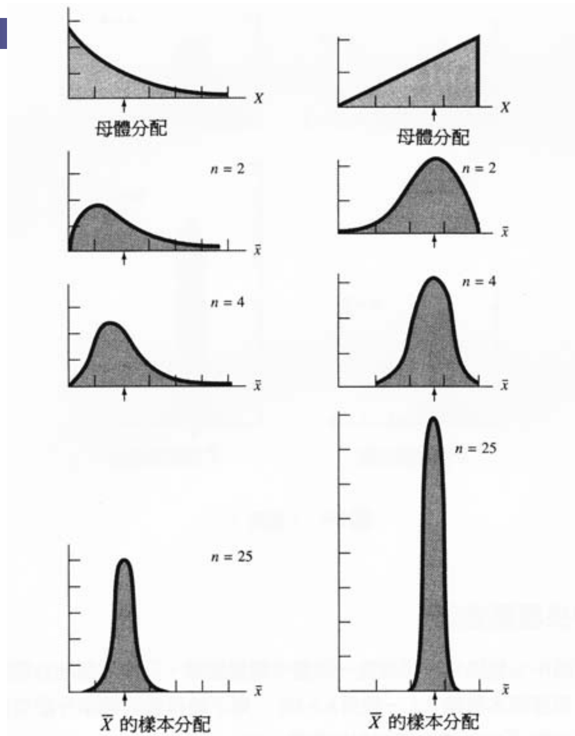

圖 9-4　中央極限定理的說明（不同 n 之 $\bar{X}$ 的抽樣分配—中央極限定理）

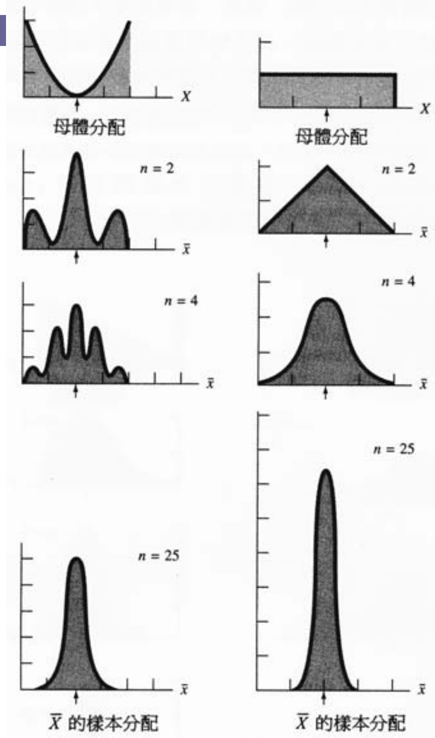

圖 9-4 （繼續）

◎當母體σ未知時

在這種情形下，我們可

(1)以與此母體特性相近的另一已知母體 σ 來代表。

(2)在大樣本(n>30)條件下，無論母體分配是否屬於常態，則以樣本標準差 S 取代之。

(3)當母體為常態分配及小樣本情形下，可用樣本S取代，惟此時須用Student-t 分配來分析。

◎自小母體中抽樣

自一有限母體中以不還原抽樣方式抽樣，則$\overline{X}$之抽樣分配近似於常態分配，而$\overline{X}$之標準誤為

$$\sigma_{\overline{X}}=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

9-4 Student-*t* 分配

(1)當抽樣為大樣本(n>30)時，不論母體是否為常態分配，其樣本平均數$\overline{X}$的分配為近似 $N(\mu,\sigma/\sqrt{n})$之常態分配。

(2)當母體為常態分配，若為小樣本(n<30)且母體標準差σ已知情形下，$\overline{X}$ 亦為常態分配。

而樣本$\overline{X}$之標準常態分配之轉換式為

$$Z=\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

(3)自母體為常態分配，但σ未知情形下抽出為小樣本(n<30)，若以樣本s估計母體σ並不合適(精確)；因此，上式並不適用。此時則須以 t 分配(與常態分配特性接近)。

◎ t 分配的統計量

(1)由於母體參數通常為未知，故其參數值另以樣本統計量來估計：即用$\overline{X}$估計μ，以 s 估計 σ。

(2)當以樣本 s 取代σ，則樣本$\overline{X}$之標準常態分配之轉換式為

$$\frac{\overline{X}-\mu}{s/\sqrt{n}}$$

由於$\overline{X}$與 s 皆為隨機變數，故經轉換後 $\overline{X}$的抽樣分配會隨著樣本數的不同而改變，故上式亦為一隨機變數。

(3)在小樣本情形下，以 s 估計 σ並不很精確。針對此問題，統計學家W.S. Gosset 在1908年以student為筆名的論文中首先提出 t 分配，或簡稱 t 分配。 **t 分配的統計量(Student Statistic)**為

$$t=\frac{\overline{X}-\mu}{s/\sqrt{n}}$$

(4)圖9-5的t 分配曲線中，α面積即為在t_α點(值)以右(右尾)的面積，其面積代表其機率值，可表為下式：

$$\alpha=\Pr[t>t_\alpha]$$

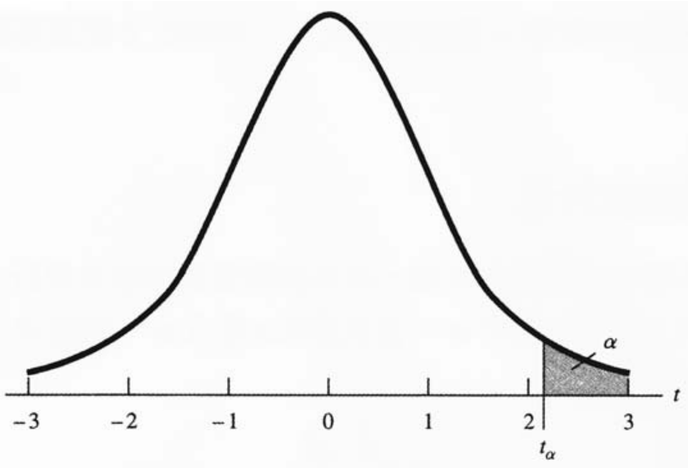

圖 9-5　t 分配之機率密度曲線 (df = 4)

◎ t 分配曲線與常態曲線之比較

t分配與常態分配相似，兩者都是對稱於0的平均值之鐘形分配。

(1)t 統計量受到兩個隨機變數$\overline{X}$與s^2的影響，而Z統計量只隨$\overline{X}$的變動而改變。

(2) Z分配的母體σ為已知而t 分配時 σ未知，故t分配的變異性較大。

(3)圖9-6可看出兩者皆為相似的鐘形曲線，但t分配的曲線的頂端較Z分配為低，且t分配曲線的尾端較Z分配為分散，表示其變異性較大。

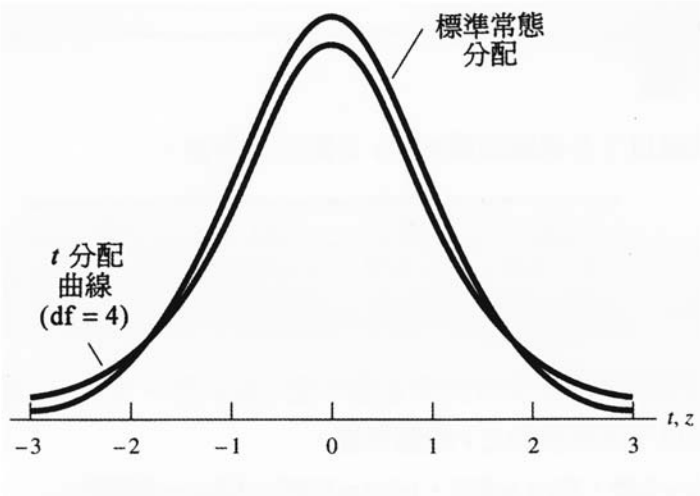

圖 9-6　標準常態密度曲線與 *t* 分配曲線之比較

9-6 樣本變異數的抽樣分配：卡方分配

欲研究未知母體**變異的程度**(Variability)，則須由樣本的s^2(或S)推論其母體的σ^2(或σ)。

◎卡方分配

◈自變異數為σ^2的母體抽出一組大小為n的隨機樣本，且其樣本的變異數為s^2，

$$X^2=\frac{(n-1)s^2}{\sigma^2}$$

自由度為$n-1$的**卡方分配**(Chi-square Distribution)，符號χ^2為卡方分配的隨機變數卡方變數。

◈卡方變數在曲線某臨界值的右尾(方)面積(機率)為α時，以χ^2_α表示，即

$$\alpha=\Pr[\chi^2>\chi^2_\alpha]$$

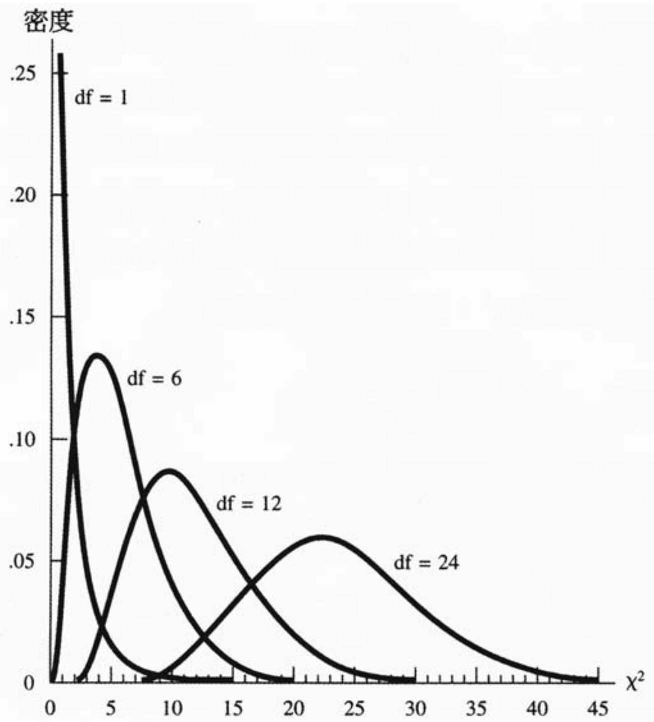

圖 9-7　不同自由度的卡方密度曲線

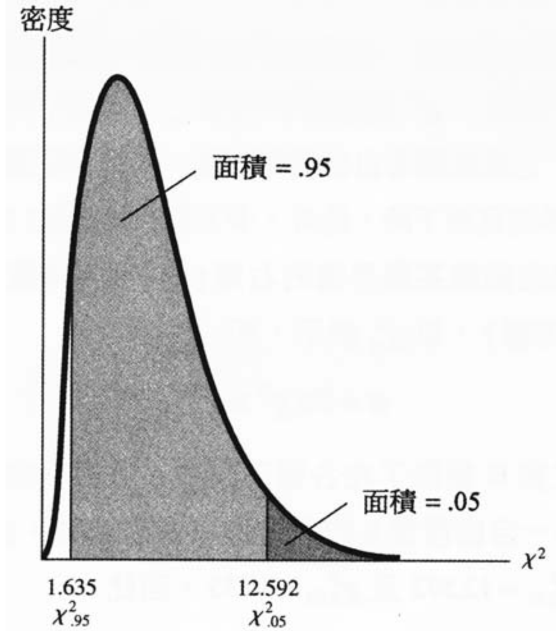

圖 9-8　自由度為 6 的卡方密度曲線

◎ **F 分配**

(1)欲對抽自兩個不同母體之兩樣本變異作比較，可求兩變異數的比率，而此變異比率分配稱為 **F 分配** (F Distrbution)。

(2)若 S_A^2 與 S_B^2 是大小為 n_A 與 n_B 的二個獨立隨機樣本的變異數，此二樣本來自變異數相同的常態母體 A 與 B，

則 F 分配可表 $$F = \frac{s_A^2}{s_B^2}$$

其自由度為 $n_A - 1$ 與 $n_B - 1$。臨界值以 F_α 表示，即

$$\alpha = \Pr[F \geq F_\alpha]$$

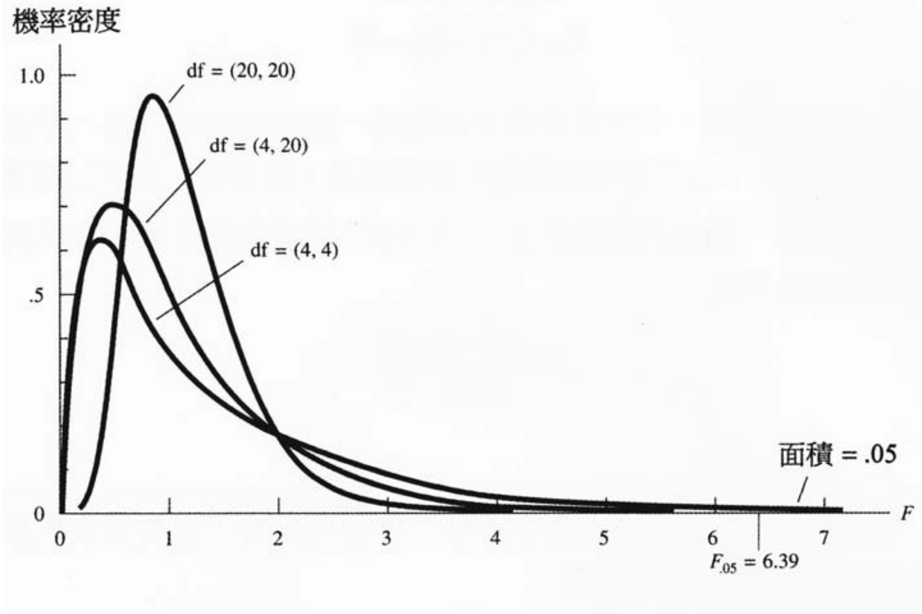

圖 9-9 不同自由度的 F 分配曲線

F 分配與 t 分配及卡方分配一樣地，均是假設母體為常態分配。