10　統計估計

- **學習目的**

1.瞭解點估計的意義、估計的步驟與限制。

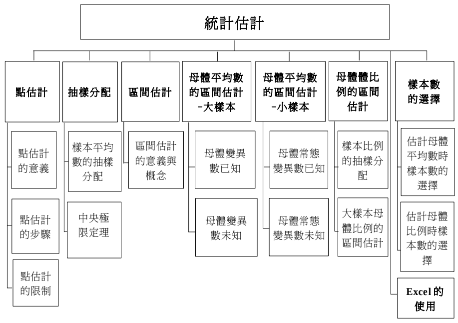
2.瞭解區間估計的意義。

3.瞭解大樣本與小樣本母體常態、變異數已知與未知下，單一母體平均數區間估計的方法。知悉*t*分配的意義與機率值。

4.瞭解單一母體比例區間估計的方法。

5.利用Excel做統計估計。

本章結構



統計估計

- **統計估計**

統計估計是利用樣本統計量去推估母體參數的方法。

點估計的意義與限制

- **點估計**

由母體抽取一組樣本數爲n的隨機樣本，並以由此得到的樣本統計量做爲母體參數的估計值。

點估計的意義與限制

- **點估計的步驟**

①抽取具代表性的樣本

②選擇一個較佳的樣本統計量做爲估計式

③計算樣本統計量的值

④以樣本統計量的值推論母體參數值並做決策

點估計的意義與限制

表10.1 台北市區房屋的價格

	A	B	C	D	E	F	G	H	I
1	700	1,520	542	840	820	854	1,280	1,490	1,388
2	856	850	928	950	1,100	960	980	956	690
3	720	790	698	1,323	848	810	998	1,050	1,080
4	835	850	821	923	753	869	860	799	930

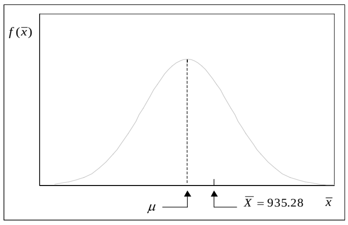
點估計的意義與限制

表10.2 台北市區的房屋價格

	A	B
1	台北市區的房屋價格	
2		
3	平均數	935.278
4	標準差	223.652
5	變異數	50020.149

點估計的意義與限制

圖10.1 台北市區房屋價格的估計



抽樣分配

- **抽樣分配**

  樣本統計量為隨機樣本的函數，而隨機樣本是由n個隨機變數$(X_1, X_2, \cdots X_n)$所組成的，故樣本統計量亦為一隨機變數，其機率分配稱為抽樣分配。

抽樣分配

樣本平均數的平均數樣

$$E(\bar{X}) = \mu_{\bar{x}} = \mu$$

樣本平均數的變異數與標準差

$$\sigma_{\bar{x}}^2 = V(\bar{X}) = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

抽樣分配

- **常態分配的加法定理**

  若母體為常態分配，則$\bar{X}$為一常態分配，即若$X \sim N(\mu, \sigma^2)$，則$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$。

抽樣分配

- **中央極限定理**

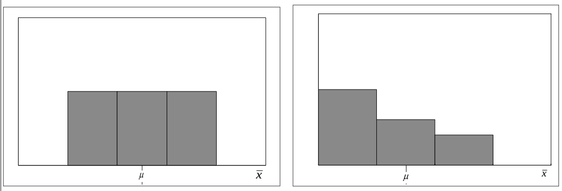
  無論母體為何種分配（常態或非常態），其平均數為μ，變異數為σ^2（$\sigma^2 < \infty$），自母體簡單隨機抽取n個元素為一組樣本，若樣本數n夠大（一般認為$n \geq 30$），則樣本平均數的抽樣分配會趨近於常態分配，即：

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

中央極限定理

圖10.2 中央極限定理

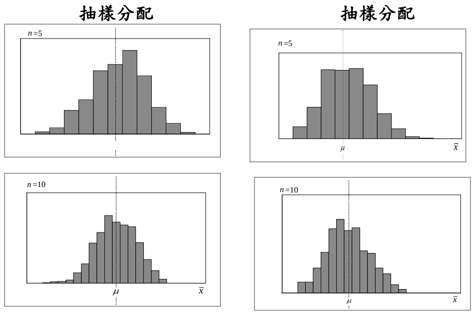
母體分配 母體分配



中央極限定理

圖10.2 中央極限定理（續）

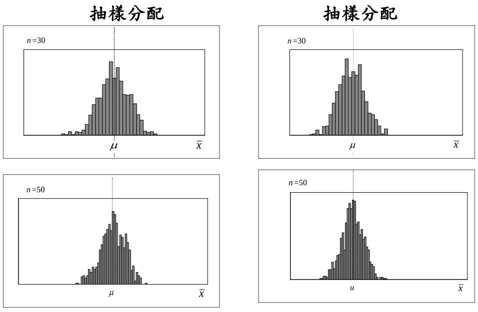
抽樣分配 抽樣分配



中央極限定理

圖10.2 中央極限定理（續）

抽樣分配 抽樣分配



區間估計的意義

- 區間估計

  對未知的母體參數估計出一個上下限的區間，並指出該區間包含母體參數的可靠度。

區間估計的意義

- 信賴區間

  信賴區間是在一個既定的信賴水準下所構成的一個區間。

  是由樣本統計量及抽樣誤差所構成的一個(包含上限，下限的)區間。

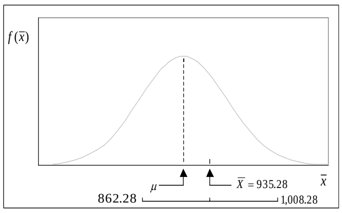
區間估計的意義

- 信賴水準(信賴係數)

  信賴水準是指信賴區間包含母體參數的信心(或稱可靠度，信賴度)。

區間估計的意義

圖10.3 台母體平均數的信賴區間



母體平均數的區間估計—大樣本

- **大樣本變異數已知，母體平均數的信賴區間**

信賴水準$1-\alpha$下，以$\bar{X}$估計μ所得的信賴區間為：

$$\bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}}$$

$\bar{X} - Z_{\alpha/2}\sigma_{\bar{X}}$ 稱為信賴區間下限，$\bar{X} + Z_{\alpha/2}\sigma_{\bar{X}}$ 稱為信賴區間上限。

- **大樣本變異數未知，母體平均數的信賴區間**

$$\bar{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

母體平均數的區間估計—大樣本

- **區間估計的步驟**

①**步驟1** 選擇較佳的點估計式並計算點估計值

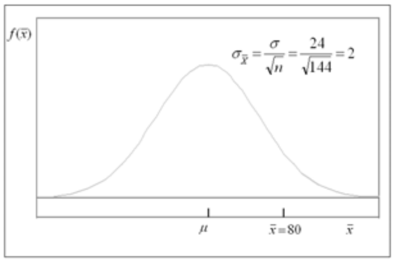
②**步驟2** 取得樣本統計量的抽樣分配

③**步驟3** 導出母體參數的信賴區間

④**步驟4** 求出母體參數的信賴區間值並做統計推論

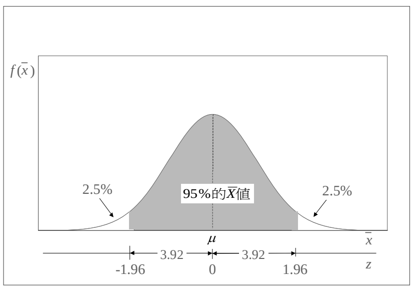
母體平均數的區間估計—大樣本

圖10.4 $\bar{X}$的抽樣分配



母體平均數的區間估計—大樣本

圖10.5 抽樣誤差小於等於1.96 $\sigma_{\bar{X}}$ 的區間



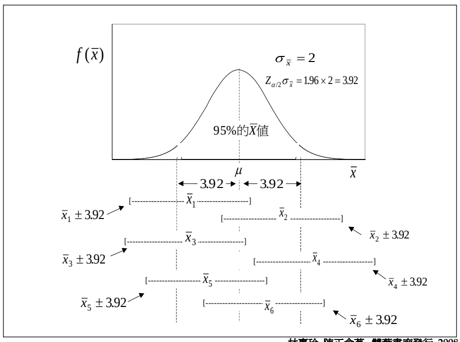
母體平均數的區間估計—大樣本

- **95%信賴水準**

95%信賴水準的含意是指，隨機抽取一組樣本所得的區間包含母體平均數的機率（或稱可靠度、信賴度）為0.95。或說區間不包含母體平均數的機率為 0.05。

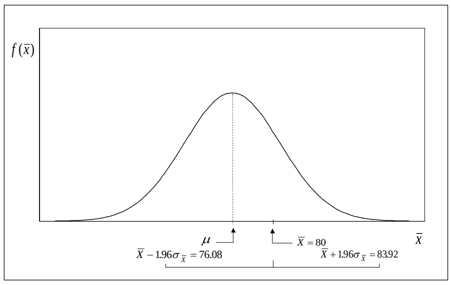
母體平均數的區間估計—大樣本

圖10.6　母體平均數 μ 的信賴區間



母體平均數的區間估計—大樣本

圖10.7　母體平均數 μ 的信賴區間



母體平均數的區間估計—大樣本

- **大樣本母體變異數已知，母體平均數的信賴區間**

  信賴水準$1-\alpha$下，以$\bar{X}$估計μ所得的信賴區間爲：

  $$\bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}}$$

  $\bar{X} - Z_{\alpha/2}\sigma_{\bar{X}}$稱爲信賴區間下限，$\bar{X} + Z_{\alpha/2}\sigma_{\bar{X}}$稱爲信賴區間上限。

母體平均數的區間估計—大樣本

表10.3 不同信賴水準下母體平均數的信賴區間

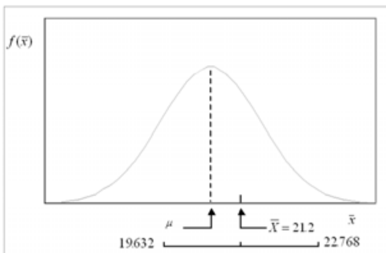
信賴水準$1-\alpha$	α	$\alpha/2$	$Z_{\alpha/2}$	信賴區間$\bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}}$
0.90	0.10	0.05	1.645	$\bar{X} \pm 1.645\sigma_{\bar{X}}$
0.95	0.05	0.025	1.96	$\bar{X} \pm 1.96\sigma_{\bar{X}}$
0.99	0.01	0.005	2.575	$\bar{X} \pm 2.575\sigma_{\bar{X}}$

母體平均數的區間估計—大樣本

圖10.8 信賴區間對話方塊

母體平均數的區間估計—大樣本

圖10.9 老年人看電視的時間



母體平均數的區間估計—大樣本

❍信賴區間長度的決定因素

①標準差的大小

若標準差愈大，則區間長度愈大；標準差愈小，則區間長度愈小。

②Z值的大小

Z值會受信賴水準的影響，信賴水準越大，信賴區間長度越長。

③樣本數*n*的大小

$\overline{X}$的標準差為$\frac{\sigma}{\sqrt{n}}$，當*n*愈大時，標準差愈小，則區間長度也就愈。

但樣本數增大時，會使抽樣成本增加。

母體平均數的區間估計—大樣本

❍ 大樣本變異數未知，母體平均數的信賴區間

$$\overline{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

母體平均數的區間估計—大樣本

表10.4 台北市區房屋價格的估計

	A	B
1	台北市區房屋價格的估計	
2		
3	平均數	935.28
4	標準誤	37.28
5	標準差	223.65
6	變異數	50020.15
7	個數	36
8	信賴度(95.0%)	75.67

母體平均數的區間估計—小樣本

❍ 小樣本常態母體變異數未知，母體平均數的信賴區間

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

為自由度*n*-1的*t*分配。

母體平均數的區間估計—小樣本

❍ *t*分配

自常態母體 $X \sim N(\mu_X, \sigma_X^2)$ 隨機抽取樣本($X_1, X_2, \ldots X_n$)，則

統計量

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

為自由度 *n*-1的 *t* 分配。

母體平均數的區間估計—小樣本

❍ *t*分配的性質

①*t* 分配為一個以平均數 0 為中心的對稱分配，不同的自由度 *v* 有不同的 *t* 分配。

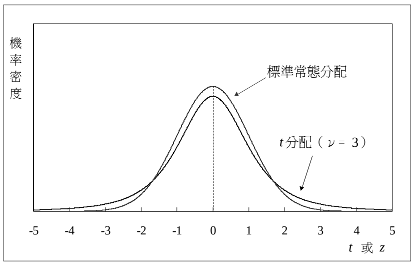
②*t* 分配不與橫軸相交。分配曲線下的總面積等於 1。*t* 分配曲線比標準常態曲線為平坦，亦即 *t* 分配曲線的高度較低，分散程度較大。

③*t* 分配的平均數與變異數為： $E(t)=0$ 、$V(t)=\frac{v}{v-2}$

④自由度趨近於無窮大時($v \to \infty$)，*t* 分配趨近於標準常態分配，即$t_v \sim N(0,1)$。一般若$v \geq 30$，則以標準常態分配代替 *t* 分配。

母體平均數的區間估計—小樣本

圖10.10　標準常態分配與t分配



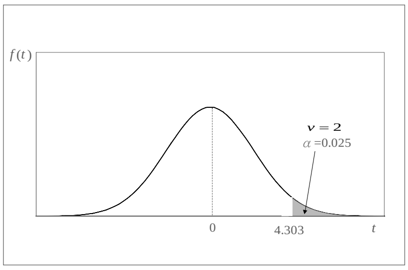
母體平均數的區間估計—小樣本

表10.5　t 值表（部份）

Df	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	10.925	2
3	1.638	2.353	3.182	4.541	5.841	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.575	∞

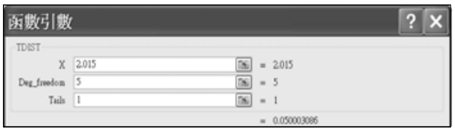
母體平均數的區間估計—小樣本

圖10.11　t分配的機率值



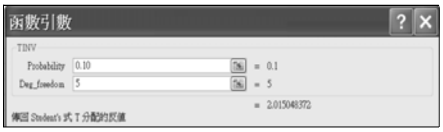
母體平均數的區間估計—小樣本

圖10.12 t 分配的對話方塊



母體平均數的區間估計—小樣本

圖10.13 t 值的對話方塊



母體平均數的區間估計—小樣本

- 小樣本常態母體變異數未知，母體平均數的信賴區間

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

母體比例的區間估計

- 樣本比例的平均數

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

母體比例的區間估計

- 樣本比例的變異數與標準差

$$V(\hat{p}) = \sigma_{\hat{p}}^2 = \frac{pq}{n}$$

$$\sigma_{\hat{p}} = \sqrt{\sigma_{\hat{p}}^2} = \sqrt{\frac{pq}{n}}$$

母體比例的區間估計

表10.6 通過時間的敘述統計

	A	B
1	通行時間	
2		
3	平均數	102.2222
4	標準誤	3.4310
5	標準差	10.2929
6	變異數	105.9444
7	個數	9
8	信賴度(95.0%)	7.9118

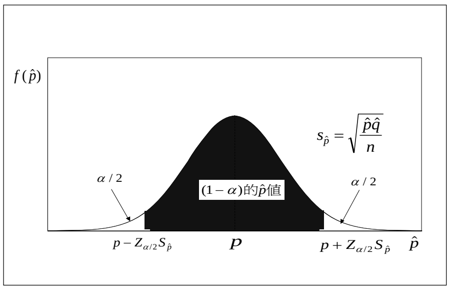
母體比例的區間估計

- 母體比例 p 的信賴區間

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

母體比例的區間估計

圖10.14 $(1-\alpha)\,\hat{p}$ 的機率區間



母體比例的區間估計

- 大樣本母體比例的信賴區間

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

母體比例的區間估計

- 大樣本母體比例的信賴區間（常用公式）

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\frac{1}{2}\times\frac{1}{2}}{n}}$$

第10章 統計估計
基礎統計學 二版

樣本數的選擇—估計母體平均數時

- 估計誤差不超過d值

$$\bar{X} - \mu = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le d$$

- 估計母體平均數時的樣本數

$$n \ge \frac{Z_{\alpha/2}^2\sigma^2}{d^2}$$

- 估計母體平均數時的樣本數(母體變異數未知)

$$n \ge \frac{Z_{\alpha/2}^2 S^2}{d^2}$$

林惠玲 陳正倉著 雙葉書廊發行 2008

第10章 統計估計
基礎統計學 二版

樣本數的選擇—估計母體比例時

- 估計母體比例時的樣本數

$$n \ge \frac{Z_{\alpha/2}^2 \hat{p}\hat{q}}{d^2}$$

- 估計母體比例時的樣本數(保守估計)

$$n \ge \frac{Z_{\alpha/2}^2(0.25)}{d^2}$$

林惠玲 陳正倉著 雙葉書廊發行 2008