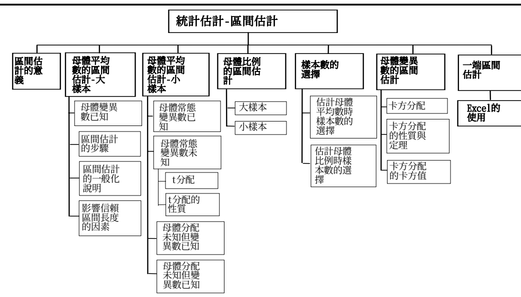
11 統計估計- 區間估計

學習目的

1.了解區間估計的意義。
2.了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。
3.了解單一母體比例區間估計的方法。
4.了解單一母體變異數區間估計的方法。了解卡方分配的意義與卡方值。
5.了解區間估計的方法在經濟、政治、社會及企業管理方面的應用。
6.利用 Excel做統計估計。

本章結構



區間估計

❍ **區間估計的意義**

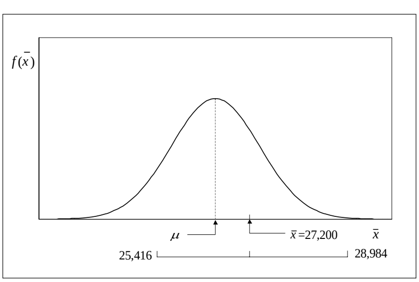
對未知的母體參數估計出一個上下限的區間，並指出該區間包含母體參數的可靠度。

信賴區間

❍ **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。

圖11.1 母體平均數的信賴區間



母體平均數的區間估計-大樣本(母體變異數已知)

❍ **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

圖11.2 $\overline{X}$ 的抽樣分配

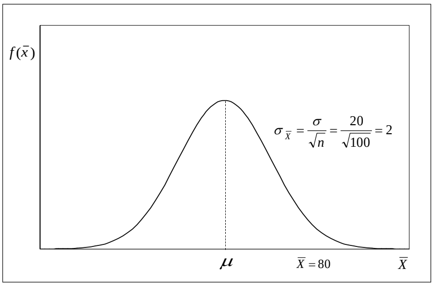
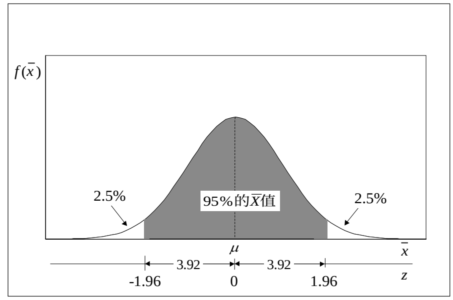


圖11.3 抽樣誤差小於等於$1.96\sigma_{\overline{X}}$的區間



信賴水準

- **信賴水準(信賴係數)**

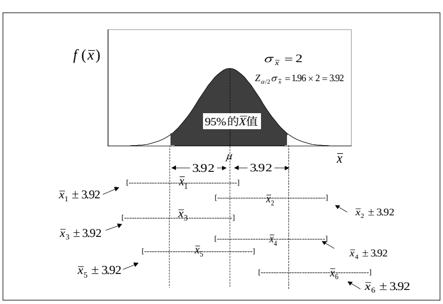
  信賴水準是指信賴區間包含母體參數的信心(或稱可靠度, 信賴度)。

信賴水準

- **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。

圖11.4 母體平均數 μ 的信賴區間



母體平均數的區間估計—大樣本

- **大樣本變異數已知，母體平均數的信賴區間**

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

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

  $\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}}$稱為信賴區間下限，$\overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$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**

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

信賴區間長度的決定因素

①所選擇的點估計式的抽樣分配

若抽樣分配的標準差越大，則信賴區間長度越長，越不精確。

②樣本數

$\overline{X}$ 的標準差為 $\sigma_{\overline{X}} = \sigma / \sqrt{n}$，當$n$ 愈大時，標準差 $\sigma_{\overline{X}}$ 愈小，區間長度也就愈小。

③機率區間上下限的取法

若上下限的取法不同，則會影響區間長度。

④信賴係數

信賴係數越大其區間長度將越大。

表11.3　小樣本的區間估計

小樣本
- 母體分配已知為常態分配
  - 母體變異數已知
  - 母體變異數未知
- 母體分配未知
  - 母體變異數已知
  - 母體變異數未知

母體平均數的區間估計—小樣本

- **小樣本常態母體變異數已知，母體平均數的信賴區間**

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

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。

*t*分配

- ***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

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

*t*分配的性質

- ***t*分配的性質**

①t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。t分配曲線下的總面積等於1。

③ t分配決定於自由度v，它是t分配唯一的參數。

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。

*t*分配

- *t* 分配的定理

  設 $X \sim N(\mu, \frac{\sigma^2}{n})$，$U \sim \chi^2_v$，$U$ 是自由度為 v 的卡方隨機變數，且 X 與 U 獨立，則隨機變數

$$\frac{(X-\mu)/\sigma}{\sqrt{U/v}} = \frac{(X-\mu)/\sigma}{\sqrt{\chi^2_v/v}} = \frac{Z}{\sqrt{\chi^2_v/v}}$$

  為自由度為 v 的 t 分配。

圖11.8 *t*分配為以0為中心的對稱分配

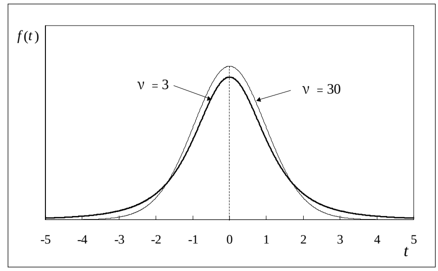


圖11.9 標準常態分配與*t*分配

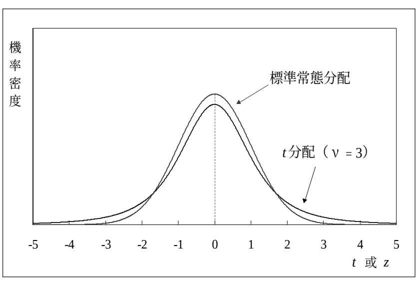


圖11.10 *t*分配的機率值

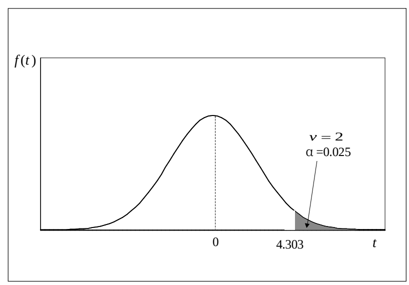
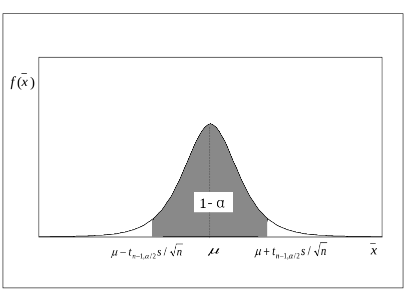


圖11.12 $(1-\alpha)\,\overline{X}$的機率區間

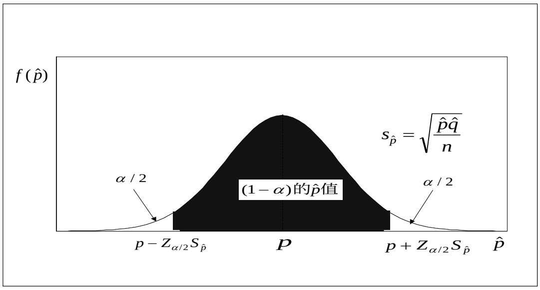


母體比例的區間估計

- 大樣本母體比例的信賴區間

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

圖11.13 $(1-\alpha)\hat{p}$的機率區間



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)

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

樣本數的選擇—估計母體比例時

- 估計母體比例時的樣本數

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

χ^2分配

- 樣本變異數

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

- 卡方統計量

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

該統計量為自由度$(n-1)$的卡方分配

χ^2分配的性質

- 卡方分配的性質

①卡方分配為一定義在大於等於0(正數)範圍的右偏分配,不同的自由度決定不同的卡方分配。

②卡方分配只有一個參數即自由度，表為*v*。卡方分配的平均數與變異數為：

$$E(\chi_v^2) = v,\quad V(\chi_v^2) = 2v$$

③卡方分配當自由度增加而逐漸對稱，當自由度趨近於無窮大時$(v \to \infty)$，卡方分配會趨近於常態分配。

④卡方分配的加法定理

二個獨立的卡方隨機變數相加所得之隨機變數仍為卡方隨機變數，其卡方分配的自由度為二個卡方分配的自由度之和。

圖11.15 卡方分配

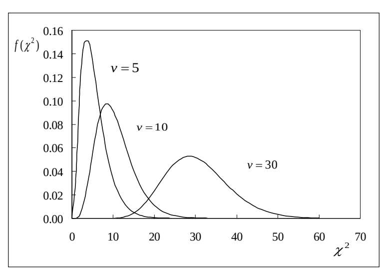
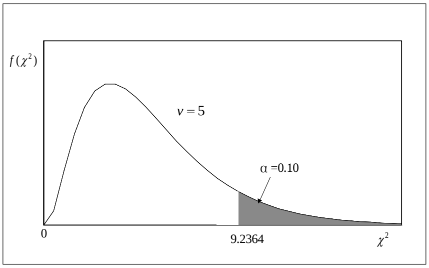


圖11.16 卡方分配機率值

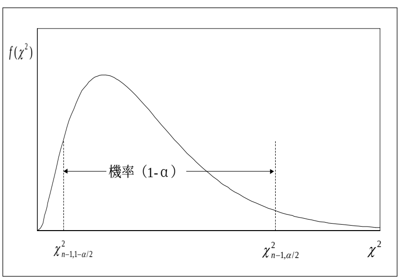


母體變異數的區間估計

- 母體變異數的信賴區間

$$\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}$$

圖11.18 $(1-\alpha)\chi^2$ 值的機率區間



一端區間估計

- 母體平均數$1-\alpha$最高限值的信賴區間

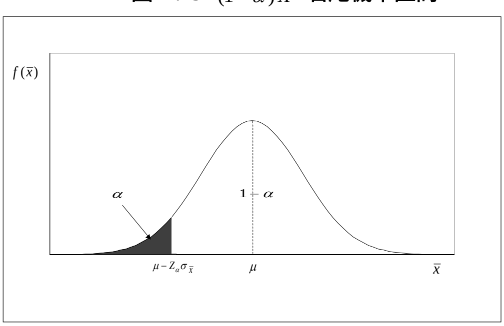
母體平均數$1-\alpha$最高限值的信賴區間為：

$$\mu \le \overline{X} + Z_\alpha \sigma_{\overline{x}}$$

或

$$(-\infty, \overline{X} + Z_\alpha \sigma_{\overline{x}})$$

圖11.19 $(1-\alpha)\overline{X}$ 右尾機率區間



一端區間估計

- 母體均數$1-\alpha$最低限值的信賴區間

$\mu \ge \overline{X} - Z_\alpha \sigma_{\overline{x}}$　或表為　$(\overline{X} - Z_\alpha \sigma_{\overline{x}}, \infty)$