第八章　估計

8.1 估計的意義

估計 $\begin{cases} \text{點估計(point estimation)} \\ \text{區間估計(interval estimation)} \end{cases}$

點估計：由抽樣所獲得的樣本

$(X_1, X_2, \ldots, X_n)$組成的統計量，

用來估計未知參數

例如：$T_1(X_1, X_2, \ldots, X_n) = \bar{X}$ 用來估計 μ

$T_2(X_1, X_2, \ldots, X_n) = S^2$ 用來估計 σ^2

或 $T_3(X_1, X_2, \ldots, X_n) = \hat{P}$ 用來估計 P

此時，統計量 $\bar{X}, S^2$ 及 $\hat{P}$ 又稱為估計量

(estimator)，將抽樣所獲得的樣本數值

代入估計量，所得到的數值稱為估計值

(estimate)

8.2 點估計量性質

- 如何求點估計量

  1.動差法(method of moments)

  2.最大概數估計法(method of likelihood estimation)

- 如何評估點估計量優劣

  1.不偏性(unbiased)

  2.有效性(efficiency)

  3.一致性(consistent)

  4.充分性(sufficiency)

  5.完全性(completeness)

- 不偏性

已知 $\mu(X_1, X_2, \ldots, X_n)$ 為未知參數 θ 之點估計量，若 E[$\mu(X_1, X_2, \ldots, X_n)$]= θ，則 $\mu(X_1, X_2, \ldots, X_n)$ 為 θ 之不偏估計量(unbiased estimator)

Ex. E($\bar{X}$)= μ

E(S^2)=σ^2

⇨ $\bar{X}, S^2$ 分別為 μ , σ^2 之不偏估計量

- 有效性

考慮母體參數 θ 之所有不偏估計量中，若 $\hat{\theta}$ 為母體參數 θ 之最小變異數不偏估計量(minimum variance unbiased estimator , MVUE)

- 相對有效性(relative efficiency)

已知 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 皆為母體參數 θ 之不偏估計量

若 $V(\hat{\theta}_1)<V(\hat{\theta}_2)$ 則 $\hat{\theta}_1$ 較 $\hat{\theta}_2$ 相對有效

$V(\hat{\theta}_1)>V(\hat{\theta}_2)$ 則 $\hat{\theta}_2$ 較 $\hat{\theta}_1$ 相對有效

- 標準誤(standard error)

估計量$\hat{\theta}$之標準誤定義為

$\sigma_{\hat{\theta}}=\sqrt{v(\hat{\theta})}$，若上述的標準誤含有未知參數，須以估計值代入，稱為被估計的標準誤，以$s_{\hat{\theta}}$或$\hat{\sigma}_{\hat{\theta}}$表示

8.3 區間估計

區間估計是以點估計量的抽樣分配為基礎，求出區間上下界，用以估計母體參數落於該區間的機率大小，即$\hat{\theta}$為母體參數θ之估計量，經由數字推導可以獲得

$$1-\alpha = P(\ell(\hat{\theta}) \le \theta \le \mu(\hat{\theta}))$$

我們稱$[\ell(\hat{\theta}), \mu(\hat{\theta})]$為未知參數$\theta$之$100(1-\alpha)\%$ 信賴區間(confidence interval 簡稱 C.I.)

$1-\alpha$：信賴水準

$\ell(\hat{\theta})$：信賴下限

$\mu(\hat{\theta})$：信賴上限

- 求信賴區間三步驟

1. 先決定統計量及抽樣分配
2. 計算出機率區間
3. 再轉換成信賴區間

8.4 母體平均數μ之區間估計

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$，$\sigma^2$已知,

求μ之$100(1-\alpha)\%$ C. I.

步驟 1　$\bar{X} \xrightarrow{推論} \mu$

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

$$\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

步驟 2

$$1-\alpha = P\left(-Z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right)$$

$$= P\left(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

則 $[\overline{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$ 為 μ 之 $100(1-\alpha)\%$ C. I.

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$，$\sigma^2$已知,

且 $n<30$, 求 μ 之 $100(1-\alpha)\%$ C. I.

步驟 1 $\bar{X} \xrightarrow{\text{推論}} \mu$

$\because \sigma^2$ 未知且 $n<30$(小樣本)

則 $\dfrac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$

步驟 2

$$1-\alpha = P\left(-t_{\alpha/2}(n-1) < \frac{\bar{X}-\mu}{s/\sqrt{n}} < t_{\alpha/2}(n-1)\right)$$

$$=P\left(-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right.$$

$$< \bar{X} - \mu < t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}})$$

$$=P(-\bar{X} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

$$< -\mu < -\bar{X} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}})$$

$$=P(\bar{X} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

$$< \mu < \bar{X} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}})$$

則 $[\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},$

$\bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}]$ 為 μ 之 $100(1-\alpha)\%$ C. I.

- μ之區間估計整理

1.母體為常態分配

Case1.σ^2已知

則$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$為$\mu$之 $100(1-\alpha)\%$ C. I.

Case2. σ^2未知,$n \geq 30$

則$\left[\bar{x} - Z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$為$\mu$之近似 $100(1-\alpha)\%$ C. I.

Case3. σ^2未知,$n<30$

則$\left[\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right]$為$\mu$之 $100(1-\alpha)\%$ C. I.

2.母體非常態分配

Case1.σ^2已知, $n \geq 30$

則$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$為$\mu$之近似 $100(1-\alpha)\%$ C. I.

Case2. σ^2未知,$n \geq 30$

則$\left[\bar{x} - Z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$為$\mu$之近似 $100(1-\alpha)\%$ C. I.

- 何謂信賴水準為 1-α 之信賴區間

例：使用電腦模擬

從常態母體 N(6,34.52)

抽出 12 組樣本數為 n=15

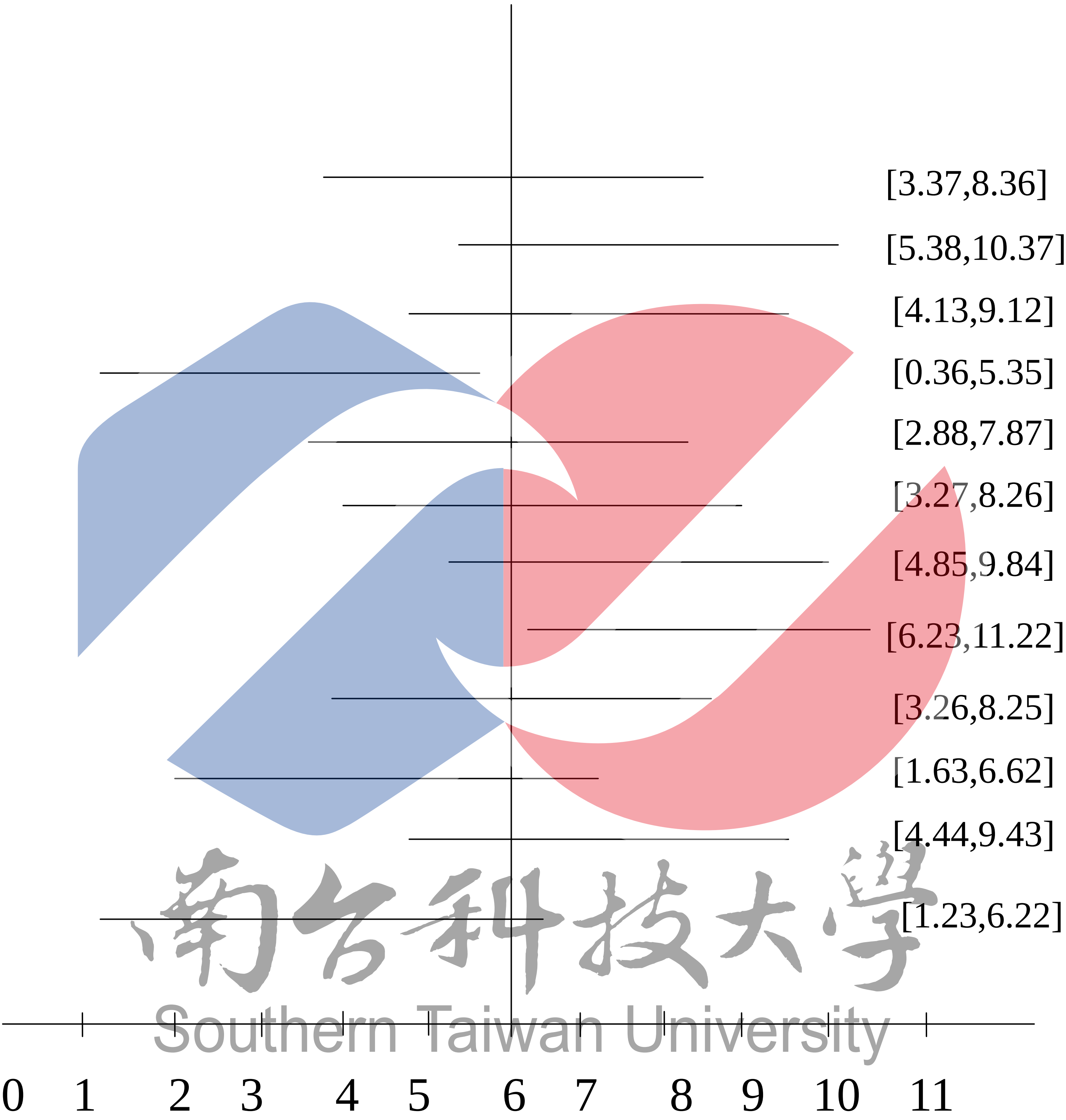
求 μ 之 90% C. I.

利用公式可獲得

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\Rightarrow \left[\bar{x} - Z_{0.05}\frac{\sqrt{34.52}}{\sqrt{15}}, \bar{x} + Z_{0.05}\frac{\sqrt{34.52}}{\sqrt{15}}\right]$$

$$\Rightarrow \left[\bar{x} - 2.4955, \bar{x} + 2.4955\right]$$



μ之90% C.I.

⇨ 第 4,8 組信賴區間不包含 6

8.5 自由度(degree of freedom)

令X_1, X_2, X_3, X_4, X_5為獨立隨機變數

則①$X_1 + X_2 + X_3 + X_4 + X_5$之自由度為 5

②$X_1 + X_2 + X_3 + X_4 + X_5$=20 之自由度為 4

同理，計算$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$時，必須先知道$\bar{X}$值，因此就只有 n-1 個樣本值可自由變動，因此自由度為 n-1。

8.7 影響信賴區間精確度的因素

信賴區間長度=上信賴界限-下信賴界限

通常希望信賴區間長度愈小，而信賴水準愈大

- 點估計量
- 樣本數
- 信賴界限
- 信賴水準

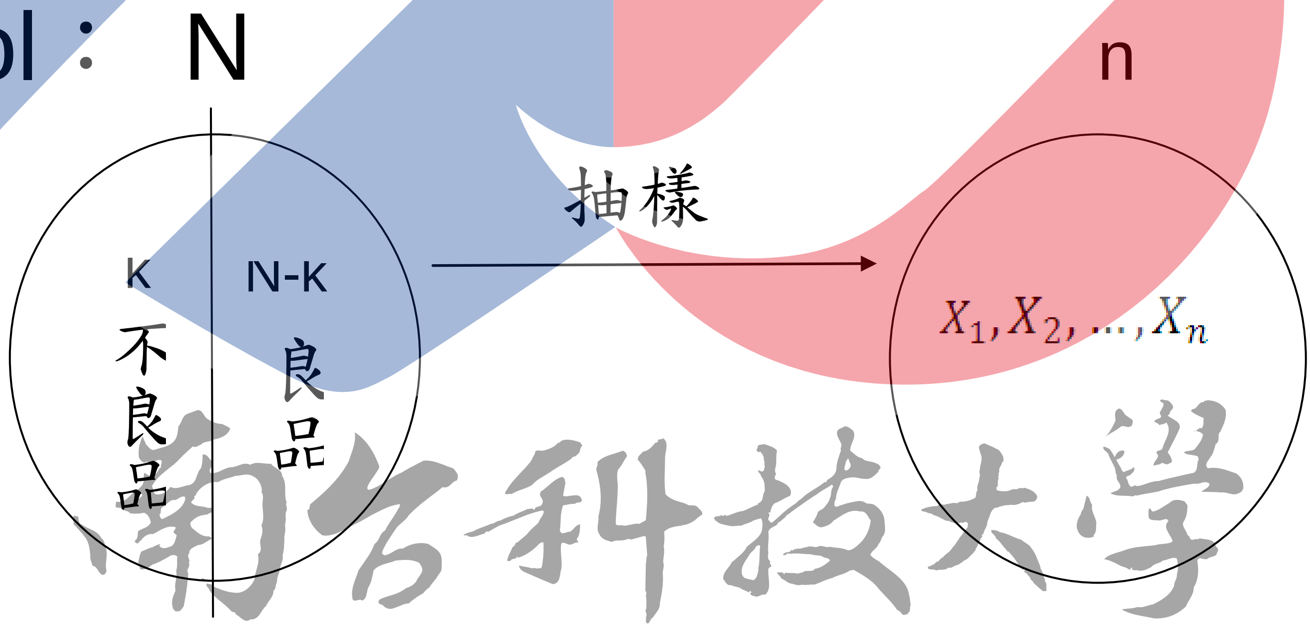
8.8 母體比例 P 之區間估計

已知一生產線共生產 N 個產品(母體)，其中 P 表示母體不良率，令隨機檢驗 n 個產品，求

①P 的點估計

②P 之 $100(1-\alpha)\%$ C. I.

Sol：



母體不良率 $P=\dfrac{k}{N}$(未知)

令 $X_i=\begin{cases} 1, & \text{表示第 i 個產品檢驗為不良品} \\ 0, & \text{表示第 i 個產品檢驗為良品} \end{cases}$

則樣本不良率$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$為母體不良率之點估計

令 $Y=\sum_{i=1}^{n} X_i$表示 n 個檢驗產品中不良品個數

假設 N 夠大，使得$\frac{n}{N} \leq 0.05$

則 Y~B(n,P)

且 E(Y)=nP

V(Y)=nP(1-P)

令 n≥30(大樣本)

由中央極限定理

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[\text{近似}]{} N(P, \frac{P(1-P)}{n})$$

即 $\dfrac{\hat{p}-P}{\sqrt{\dfrac{P(1-P)}{n}}}$ $\widetilde{\text{近似}}$ N(0,1)

$$1-\alpha \approx P\left(-Z_{\alpha/2}<\frac{\hat{P}-P}{\sqrt{\frac{P(1-P)}{n}}}<Z_{\alpha/2}\right)$$

$$= P\left(\hat{P}-Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}<P<\hat{P}+Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}\right)$$

以 $\hat{p}$ 估計 P

則

$$\left[\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

為 P 之近似 100(1-α)% C. I.