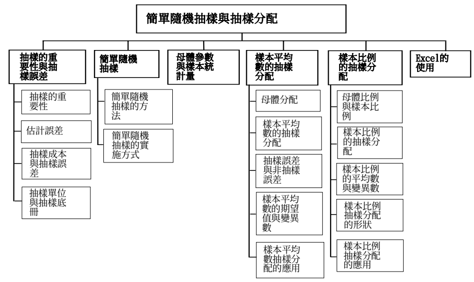
9 簡單隨機抽樣與抽樣分配

學習目的

1. 了解抽樣的意義以及為什麼要抽樣。
2. 了解機率抽樣與非機率抽樣及其優缺點與使用時機。
3. 知悉樣本大小、抽樣成本和抽樣誤差的關係。
4. 了解樣本統計量：樣本平均數、樣本比例的抽樣分配的形狀及其平均數、變異數的計算。
5. 了解中央極限定理及其應用。
6．利用Excel來做抽樣。

本章結構



抽樣的重要性

統計推論係利用樣本統計量去推論母體的特質，而樣本是否具有代表性會受抽樣方法的影響，因此抽樣方法非常重要。

估計誤差

- **抽樣誤差**

  抽樣誤差是樣本統計量與相對應的母體參數間的差異。此種差異來自抽樣過程的機遇(chance)，抽樣方法及推論方法的不同。

- **非抽樣誤差**

  非抽樣誤差主要來自調查時的執行與事後在記錄、整理資料時所發生的錯誤。

圖9.4 估計誤差

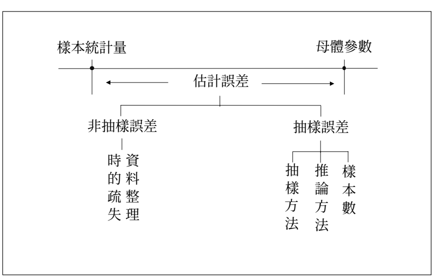
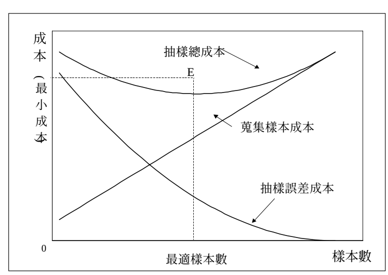


圖9.5 資料搜集成本與抽樣誤差



簡單隨機抽樣

- **簡單隨機抽樣的意義**

  抽取樣本時，若所有可能抽出的樣本被抽出的機率均相等，則稱該抽樣方法為簡單隨機抽樣。

- **簡單隨機抽樣的實施方式**

  ①抽籤式

  ②以亂數表抽取樣本

  ③以電腦做隨機抽樣

母體參數與樣本統計量

- **母體參數**

  母體參數是描述母體資料特性的統計測量數，一般簡稱為參數或母數。參數是我們想要獲取的，是統計的核心。

- **樣本統計量**

  樣本統計量為樣本的實數函數。

- **抽樣分配**

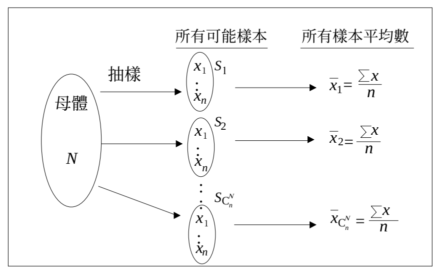
  樣本統計量為隨機樣本的函數，而隨機樣本是由n個隨機變數$(X_1, X_2, \cdots X_n)$所組成的，故樣本統計量亦為一隨機變數，其機率分配稱為抽樣分配。

樣本平均數的抽樣分配

- **樣本平均數的抽樣分配**

  設母體為隨機變數X，其機率分配為$f(x)$，若自母體中簡單隨機抽取n個元素為一組樣本，表為$(X_1, X_2, ..., X_n)$，若令$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$，則$\overline{X}$為樣本平均數。其機率分配表為$f(\overline{x})$，稱為樣本平均數的抽樣分配。

圖9.7　樣本平均數的抽樣分配



樣本平均數的期望值與變異數

- **$\overline{X}$抽樣分配的平均數與變異數**

  $\overline{X}$抽樣分配的平均數與變異數稱為$\overline{X}$的平均數與變異數。以符號$\mu_{\overline{X}}$或$E(\overline{X})$及$\sigma^2_{\overline{X}}$或$V(\overline{X})$分別表示。

- **$\overline{X}$抽樣分配的平均數**

  $\overline{X}$抽樣分配的平均數等於母體平均數，即

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

$\overline{X}$抽樣分配的變異數與標準差

- **無限母體樣本平均數的變異數($\sigma^2_{\overline{X}}$)與標準差($\sigma_{\overline{X}}$)**

  $$\sigma^2_{\overline{X}} = V(\overline{X}) = \frac{\sigma^2}{n}$$

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

- **有限母體抽出不放回樣本平均數的變異數($\sigma^2_{\overline{X}}$)與標準差($\sigma_{\overline{X}}$)**

  $$\sigma^2_{\overline{X}} = V(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

  $$\sigma_{\overline{X}} = \sqrt{V(\overline{X})} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

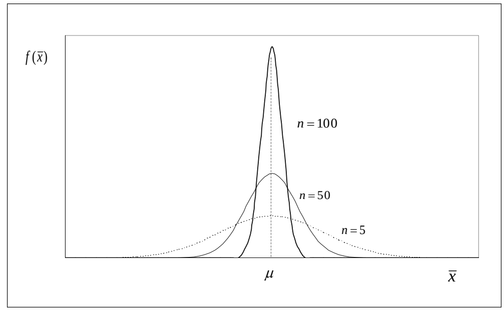
大數法則

- **大數法則**

$$\lim_{n\to\infty} P(|\bar{X}-\mu|<\varepsilon)=1$$

亦即當樣本數夠大時，$\bar{X}$ 會趨近於 μ 的機率為1。

圖9.16　大數法則



樣本平均數抽樣分配的形狀

- **常態母體 $\bar{X}$ 的抽樣分配**

若母體為常態分配，平均數為 μ，標準差為 σ，則不論樣本數為何，樣本平均數 $\bar{X}$ 的抽樣分配亦為常態分配，其平均數和標準差分別為：

$\mu_{\bar{X}}=\mu$，$\sigma_{\bar{X}}=\dfrac{\sigma}{\sqrt{n}}$

- **中央極限定理(非常態母體 $\bar{X}$ 的抽樣分配)**

無論母體為何種分配，自母體簡單隨機抽取 n 個為一組樣本，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。

樣本平均數抽樣分配的形狀

- **應用中央極限定理的注意事項**

①一般而言，不論母體為何種分配，當 $n\geq 30$ 時，$\bar{X}$ 漸趨於常態分配。

②中央極限定理可適用於樣本統計量為隨機樣本 $(X_1,X_2,\cdots,X_n)$ 線性函數的情況。

③若母體為非常態分配雖是大樣本，則其抽樣分配不是常態分配，而是近似常態分配。

④中央極限定理僅適用於大樣本。

表9.11　$\bar{X}$ 的抽樣分配

樣本	母體分配	抽樣分配
大樣本 $(n\geq 30)$	母體為常態分配	$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$
	母體非常態分配	$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$
小樣本 $(n<30)$	母體為常態分配	$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$
	母體非常態分配	$\bar{X}$ 的分配決定於母體分配

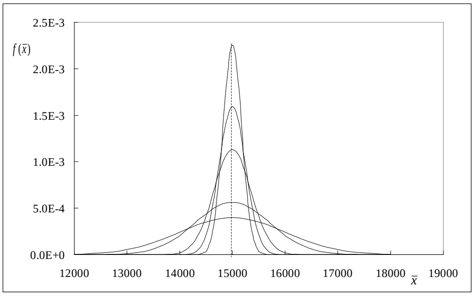
註：①若母體為有限母體，且 $\frac{n}{N}>0.05$，則 $V(\bar{X})=\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$。

②若母體為有限母體，且 $n/N>0.05$，則 $\bar{X}$ 不一定為常態分配，因 $(X_1,\cdots,X_n)$ 不獨立。

表9.12　樣本數不同的抽樣分配

樣本數	$E(\bar{X})=\mu$	$\sigma_{\bar{X}}=\sigma/\sqrt{n}$
$n=1$	15,000	1000/1=1000
$n=2$	15,000	$1000/\sqrt{2}=707.11$
$n=8$	15,000	$1000/\sqrt{8}=353.55$
$n=16$	15,000	$1000/\sqrt{16}=250.00$
$n=32$	15,000	$1000/\sqrt{32}=176.78$

圖9.20　樣本大小不同的抽樣分配



樣本比例的抽樣分配

- **母體比例**

  $p = K / N$

  N：母體個數，K：母體中A類別的個數。

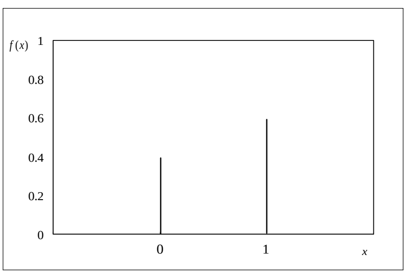
- **樣本比例**

  $$\hat{p} = \frac{k}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- **樣本比例的平均數**

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

圖9.23　點二項分配



樣本比例的抽樣分配

- **樣本比例的變異數與標準差**

  ①無限母體

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

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

  ②有限母體

  $$V(\hat{p}) = \sigma_{\hat{p}}^2 = \frac{pq}{n} \cdot \frac{N-n}{N-1}$$

  $$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}$$

表9.15　樣本比例的抽樣分配

母體為一點二項分配
$E(X) = p$
$V(X) = pq$

- 大樣本 $np>5, nq>5$
  - 母體無限　$\hat{p} \sim N(p, \frac{pq}{n})$
  - 母體有限　$\hat{p} \sim N(p, \frac{pq}{n}\frac{N\text{-}}{N\text{-}1})$
- 小樣本
  - 母體無限　$\hat{p}$ ~ 二項分配$(p, \frac{pq}{n})$
  - 母體有限　$\hat{p}$ ~ 超幾何分配$(p, \frac{pq}{n}\frac{N\text{-}n}{N\text{-}1})$

圖9.26　亂數產生器隨機抽樣對話方塊