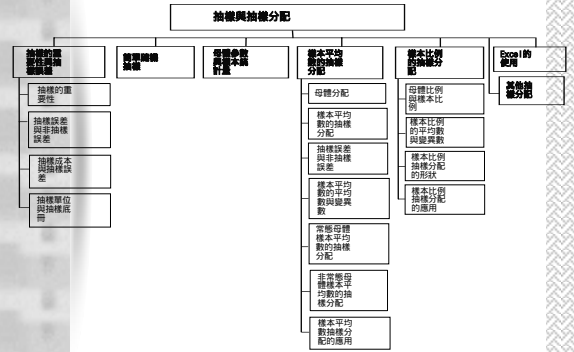


9 抽樣與抽樣分配

學習目的

1. 了解抽樣的意義以及抽樣的重要性。
2. 了解抽樣誤差與非機率誤差。
3. 知悉樣本大小、抽樣成本和抽樣誤差的關係。
4. 了解簡單隨機抽樣。
5. 了解樣本統計量：樣本平均數、樣本比例的抽樣分配的形状及其平均數、變異數的計算。
6. 了解中央極限定理及其應用。
7. 了解其他抽樣方法。
8. 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。

圖9.1 等待看牙時間(母體)

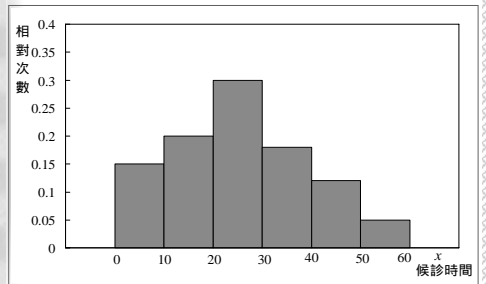


圖9.2 等待看牙時間(樣本1)

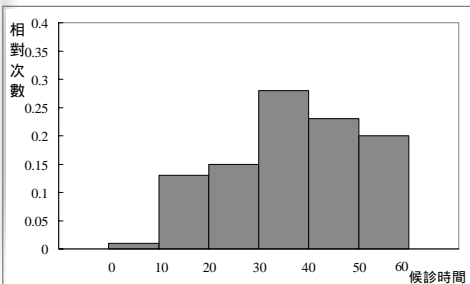
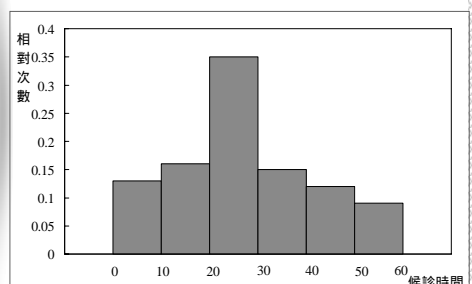


圖9.3 等待看牙時間(樣本2)



第9章 抽樣與抽樣分配 應用統計學

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。

林惠玲 陳正倉著 雙葉書報發行 2002

第9章 抽樣與抽樣分配 應用統計學

圖9.4 估計誤差

The diagram illustrates the components of estimation error. At the top, '樣本統計量' (Sample Statistics) and '母體參數' (Population Parameters) are connected by a double-headed arrow labeled '估計誤差' (Estimation Error). Below this, the error is divided into two main categories: '非抽樣誤差' (Non-sampling Error) and '抽樣誤差' (Sampling Error). '非抽樣誤差' includes '時資' (Timing), '料疏' (Data Omission), and '失理' (Misconduct). '抽樣誤差' is further divided into '抽樣方法' (Sampling Method), '推論方法' (Inference Method), and '樣本數' (Sample Size).

林惠玲 陳正倉著 雙葉書報發行 2002

第9章 抽樣與抽樣分配 應用統計學

圖9.5 資料搜集成本與抽樣誤差

The graph plots '成本 (最小成本)' (Cost) on the y-axis against '樣本數' (Sample Size) on the x-axis. Three curves are shown: '抽樣總成本' (Total Sampling Cost), '蒐集樣本成本' (Cost of Collecting Samples), and '抽樣誤差成本' (Cost of Sampling Error). '蒐集樣本成本' is a straight line starting from the origin. '抽樣誤差成本' is a downward-sloping curve. '抽樣總成本' is the sum of the other two, forming a U-shape. A vertical dashed line marks the '最適樣本數' (Optimal Sample Size) at the minimum point of the total cost curve, with a corresponding horizontal dashed line to the y-axis labeled 'E'.

林惠玲 陳正倉著 雙葉書報發行 2002

第9章 抽樣與抽樣分配 應用統計學

抽樣底冊

- **抽樣單位與抽樣底冊**
抽樣單位是抽樣母體中的一個母體元素或一組母體元素。抽樣底冊是抽樣單位的名冊或一覽表。

林惠玲 陳正倉著 雙葉書報發行 2002

第9章 抽樣與抽樣分配 應用統計學

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 用電腦做隨機抽樣

林惠玲 陳正倉著 雙葉書報發行 2002

第9章 抽樣與抽樣分配 應用統計學

表9.1 亂數表

	1	2	3	4	5	6	7	8	9	10
1	6824	7709	3937	3289	9545	0620	3904	5203	6590	8769
2	0237	7574	8607	1502	4776	0944	4946	1519	4834	2810
3	1336	8960	2192	7132	9267	4262	6070	7664	7690	3873
4	6840	3016	3991	8582	1813	0012	3781	8635	0286	3932
5	5577	7452	9477	7942	7328	0822	7876	6379	9014	6845
6	3495	3500	9497	8688	7764	0017	1221	5816	8840	8573
7	5163	5127	5955	7826	0982	3563	7783	1575	7738	9146
8	3746	5767	5137	3846	9113	3394	5172	3745	2574	5275
9	0596	6736	4273	7665	8229	6933	6510	0093	4091	4567
10	6553	4267	4071	3532	0593	3874	5368	5295	6303	2629

林惠玲 陳正倉著 雙葉書報發行 2002

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, \dots, 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, \dots, X_n) ，若令 $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ ，則 \bar{X} 為樣本平均數。其機率分配表為 $f(\bar{x})$ ，稱為樣本平均數的抽樣分配。

表9.2 展示小姐的月薪的次數分配

x	f
22	1
25	2
28	1
30	1
$N = 5$	

圖9.8 展示小姐月薪的母體機率分配

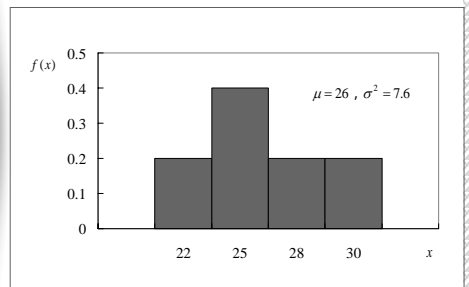


圖9.9 樣本平均數的抽樣分配

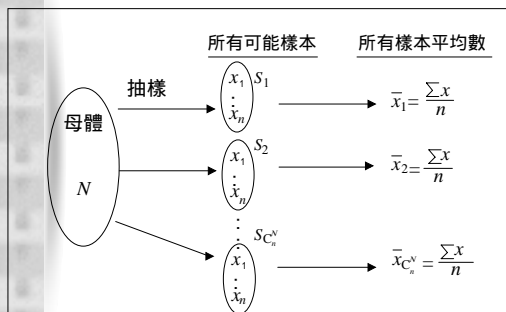
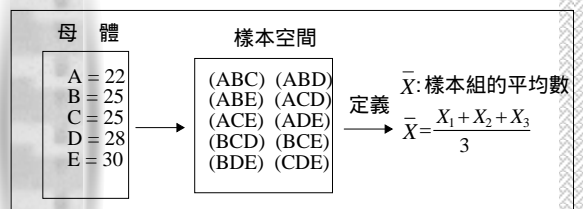


圖9.10 展示接待小姐月薪的抽樣



第9章 抽樣與抽樣分配 應用統計學

表9.5 展示小姐月薪的樣本平均數

樣本	樣本平均數 \bar{x}
(ABC) = (22,25,25)	24
(ABD) = (22,25,28)	25
(ABE) = (22,25,30)	25.6667
(ACD) = (22,25,28)	25
(ACE) = (22,25,30)	25.66667
(ADE) = (22,28,30)	26.66667
(BCD) = (25,25,28)	26
(BCE) = (25,25,30)	26.66667
(BDE) = (25,28,30)	27.66667
(CDE) = (25,28,30)	27.66667

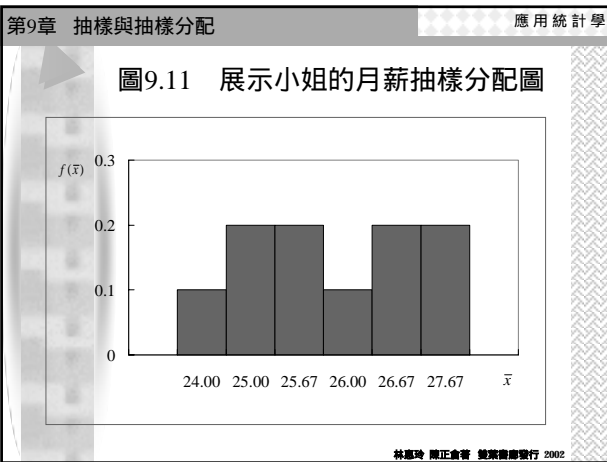
林基琦 陳正倉著 雙葉書局發行 2002

第9章 抽樣與抽樣分配 應用統計學

表9.6 展示小姐的月薪的抽樣分配

\bar{x}	\bar{x}^2	$f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
24	576	1/10 = 0.10	57.6
25	625	2/10 = 0.20	125
25.66667	659.7779489	2/10 = 0.20	131.7555898
26	676	1/10 = 0.10	67.6
26.66667	711.1112889	2/10 = 0.20	142.222578
27.66667	765.4446289	2/10 = 0.20	153.099258
$\sum f(\bar{x}) = 1.00$			677.2667733

林基琦 陳正倉著 雙葉書局發行 2002



- 第9章 抽樣與抽樣分配 應用統計學
- 樣本平均數的平均數與變異數
- \bar{X} 的平均數與變異數
 \bar{X} 抽樣分配的平均數與變異數稱為 \bar{X} 的平均數與變異數。以符號 $\mu_{\bar{X}}$ 或 $E(\bar{X})$ 及 $\sigma_{\bar{X}}^2$ 或 $V(\bar{X})$ 分別表示。
 - \bar{X} 抽樣分配的平均數
 \bar{X} 抽樣分配的平均數等於母體平均數，即

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$
- 林基琦 陳正倉著 雙葉書局發行 2002

- 第9章 抽樣與抽樣分配 應用統計學
- 樣本平均數的平均數與變異數
- 無限母體樣本平均數的變異數($\sigma_{\bar{x}}^2$)與標準差($\sigma_{\bar{x}}$)

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

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
 - 有限母體抽出不放回樣本平均數的變異數($\sigma_{\bar{x}}^2$)與標準差($\sigma_{\bar{x}}$)

$$\sigma_{\bar{x}}^2 = V(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

$$\sigma_{\bar{x}} = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$
- 林基琦 陳正倉著 雙葉書局發行 1999

第9章 抽樣與抽樣分配 應用統計學

表9.10 樣本平均數的抽樣分配

	A	B	C
1	平均數的抽樣分配		
2			
3	平均數	9.98818566	
4	標準誤	0.033788615	
5	標準差	0.337886149	
6	變異數	0.11416705	
7	範圍	1.537430535	
8	最小值	9.229333048	
9	最大值	10.76676358	
10	個數	100	

林基琦 陳正倉著 雙葉書局發行 2002

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：

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

○ 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。

圖9.20 中央極限定理
母體分配 母體分配

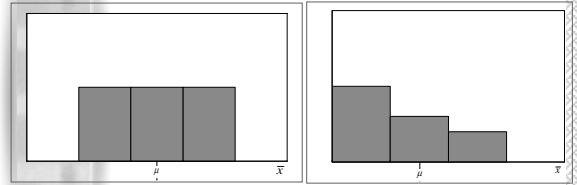


圖9.20 中央極限定理 (續)

抽樣分配 抽樣分配

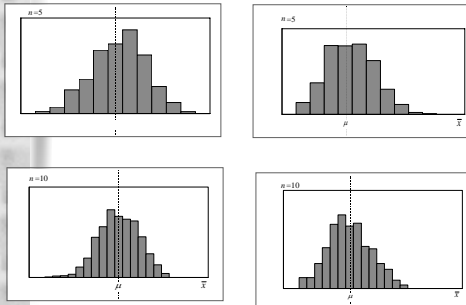
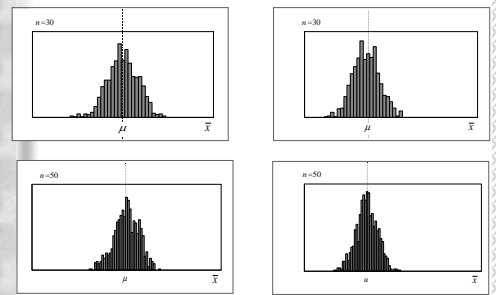


圖9.20 中央極限定理 (續)

抽樣分配 抽樣分配



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, \dots, 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。

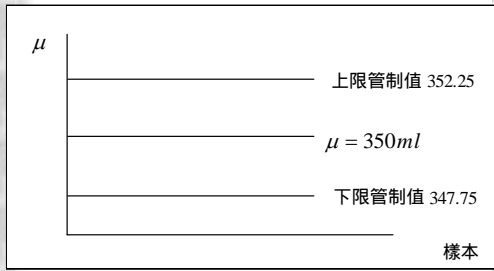
表9.8 \bar{X} 的抽樣分配

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, \dots, X_n) 不獨立。

圖9.16 品質管制圖



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。

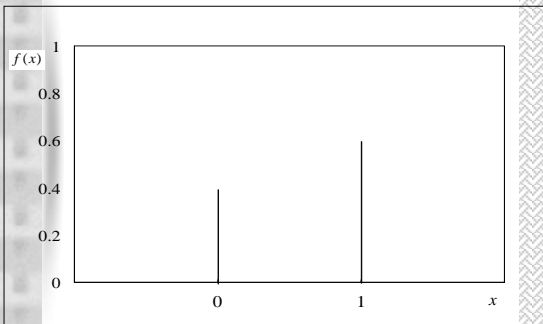
○ 樣本比例

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

○ 樣本比例的平均數

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

圖9.17 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

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

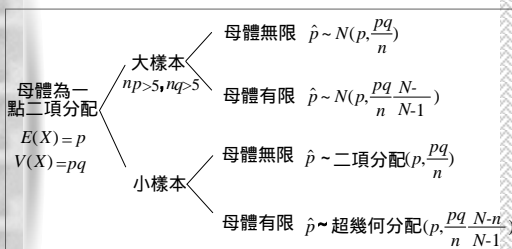
$$\sigma_p = \sqrt{\sigma_p^2} = \sqrt{\frac{pq}{n}}$$

② 有限母體

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

$$\sigma_p = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}$$

表9.10 樣本比例的抽樣分配

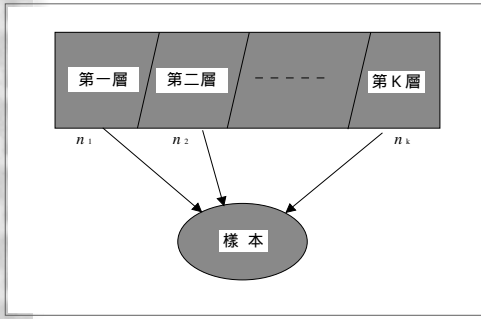


分層抽樣

○ 分層抽樣的意義

分層抽樣是將母體依其特性或依與調查目的有關的性質分成幾個類或組, 母體中的每一個個體或元素都屬於其中的一層, 而且是唯一的一層。分層之後再從各層中簡單隨機抽取樣本。

圖9.18 分層抽樣方法 - 比例抽樣

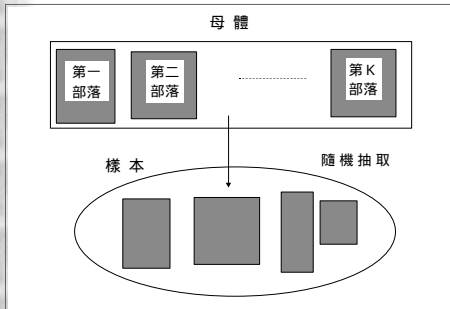


部落抽樣

○ 部落抽樣的意義

部落抽樣是先將母體中相鄰的某些群體劃分為 n 個不同的部落 (cluster)，母體中的每一個元素均屬於其中的一個部落，且是唯一的一個部落。然後再從這些部落中隨機抽取部落，並對抽出的部落進行普查的抽樣方法，又稱集團抽樣。

圖9.19 部落抽樣法

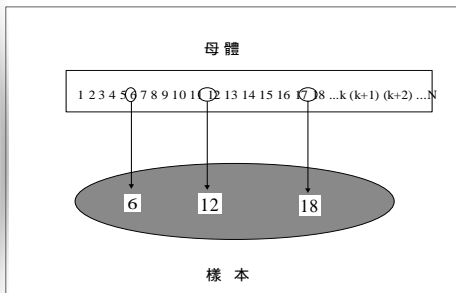


系統抽樣

○ 系統抽樣的意義

系統抽樣法是自母體自然隨機排列的資料中，每隔一定間隔選取一個樣本，直至抽滿 n 個樣本為止。

圖9.20 系統抽樣法

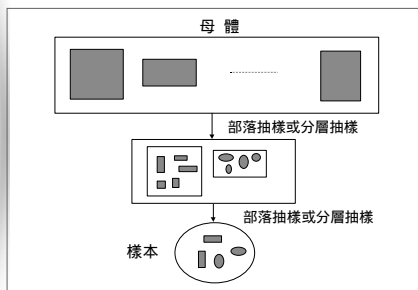


分段抽樣

○ 分段抽樣的意義

分段抽樣法是將母體按照某些特性或某種分類標準分為數個部落或層別，先由這些部落或層別中抽出幾個部落或層別，此為第一段。然後再由已經抽出的部落或層別，依特性或分類標準再抽出部落或層別，此為第二段，如此依序為之，最後再依隨機或系統或其它方法抽出樣本。

圖9.22 分段抽樣法

**非機率抽樣法**○ **判斷抽樣法**

統計人員或調查研究人員根據自己的專長、知識、研究的目的來選取代表性的樣本，此種抽樣方法稱為判斷抽樣法，又稱為目的抽樣法。

○ **方便抽樣法**

方便抽樣法是調查研究人員以現有的或方便取得的方式來抽取樣本的方法。