第 9 章　簡單隨機抽樣與抽樣分配

林惠玲　陳正倉著 雙葉書廊發行 2004

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

- 簡單隨機抽樣與抽樣分配
  - 抽樣的重要性與抽樣誤差
    - 抽樣的重要性
    - 抽樣誤差與非抽樣誤差
    - 抽樣成本與抽樣誤差
    - 抽樣單位與抽樣底冊
  - 簡單隨機抽樣的方法
    - 簡單隨機抽樣的方法
    - 簡單隨機抽樣的實施方法
  - 母體參數與樣本統計量
  - 樣本平均數的抽樣分配
    - 母體分配
    - 樣本平均數的抽樣分配
    - 抽樣誤差與非抽樣誤差
    - 母體分配抽取方式與樣本平均數的抽樣分配
  - 中央極限定理
    - 樣本平均數的平均數與變異數
    - 樣本平均數的平均數與變異數的特性
    - 抽樣分配與樣本大小的關係
    - 常態母體樣本平均數的抽樣分配
  - 樣本平均數抽樣分配的應用
  - 樣本比例的抽樣分配
    - 母體比例與樣本比例
    - 樣本比例的抽樣分配
    - 樣本比例的平均數與變異數
    - 樣本比例抽樣分配的情形
    - 樣本比例抽樣分配的應用

林惠玲 陳正倉著 雙葉書廊發行 2004

- Concept

進行推論統計的三項重要因素

1. 樣本大小 (與抽樣成本有關)

2. 抽樣方法 (本章討論)

3. 推論方法 (第10章開始)

林惠玲 陳正倉著 雙葉書廊發行 2004

9.1 抽樣的重要性與抽樣誤差

9.1.1 抽樣的重要性

EX:

1. 行政院主計處 「台灣地區人力資源調查」
2. 2008年總統大選的民意調查
3. 進口煤碳的抽樣燃燒檢驗

…

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。

→ 抽樣方法足以影響樣本的代表性, 進而導致不同的結果, 因此對抽樣方法必須講究.

Q: 等待太久會不耐煩而離去

圖9.1　等待看牙時間（母體)

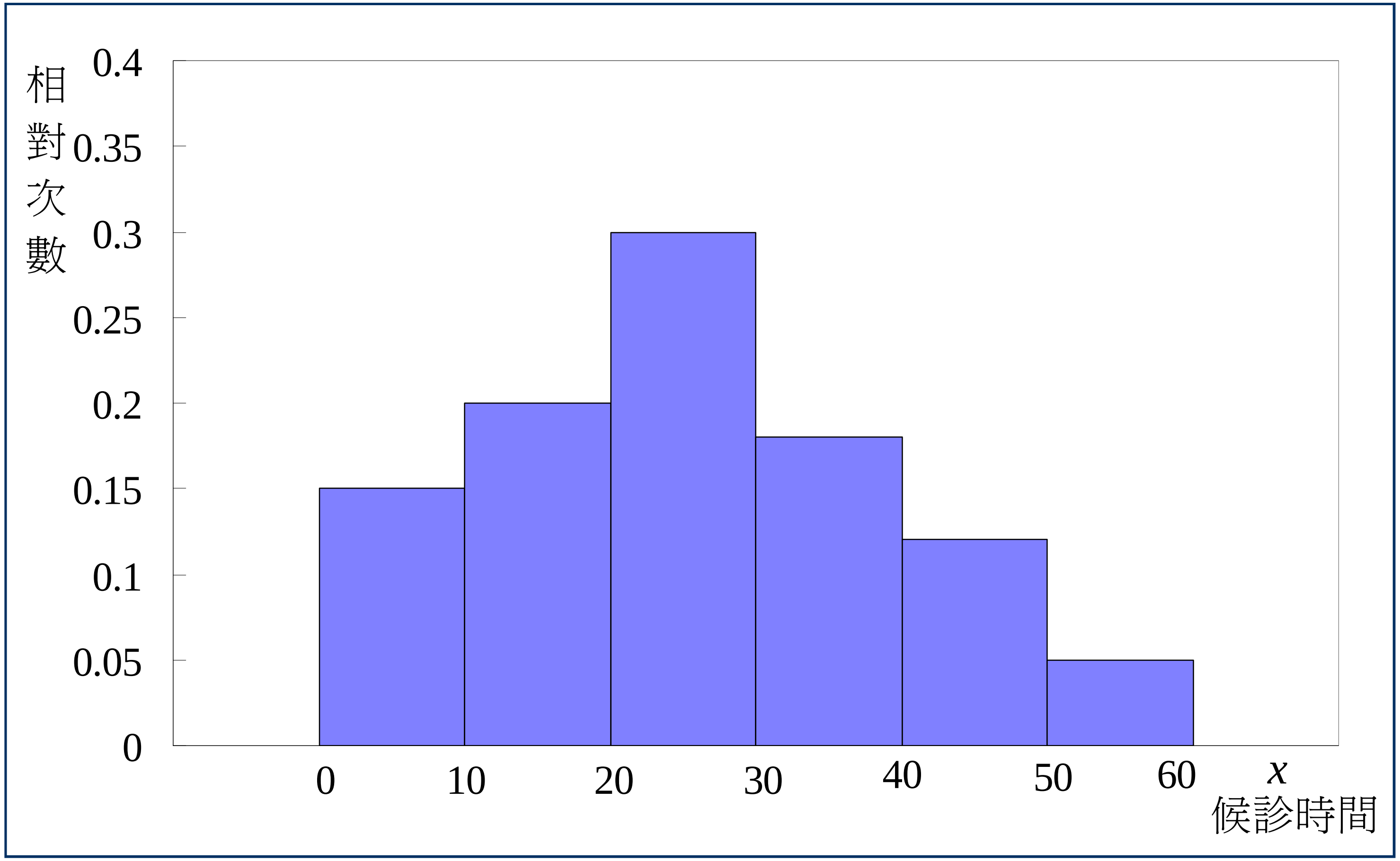


圖9.2　等待看牙時間(樣本1)

超過50分鐘的患者超過20%

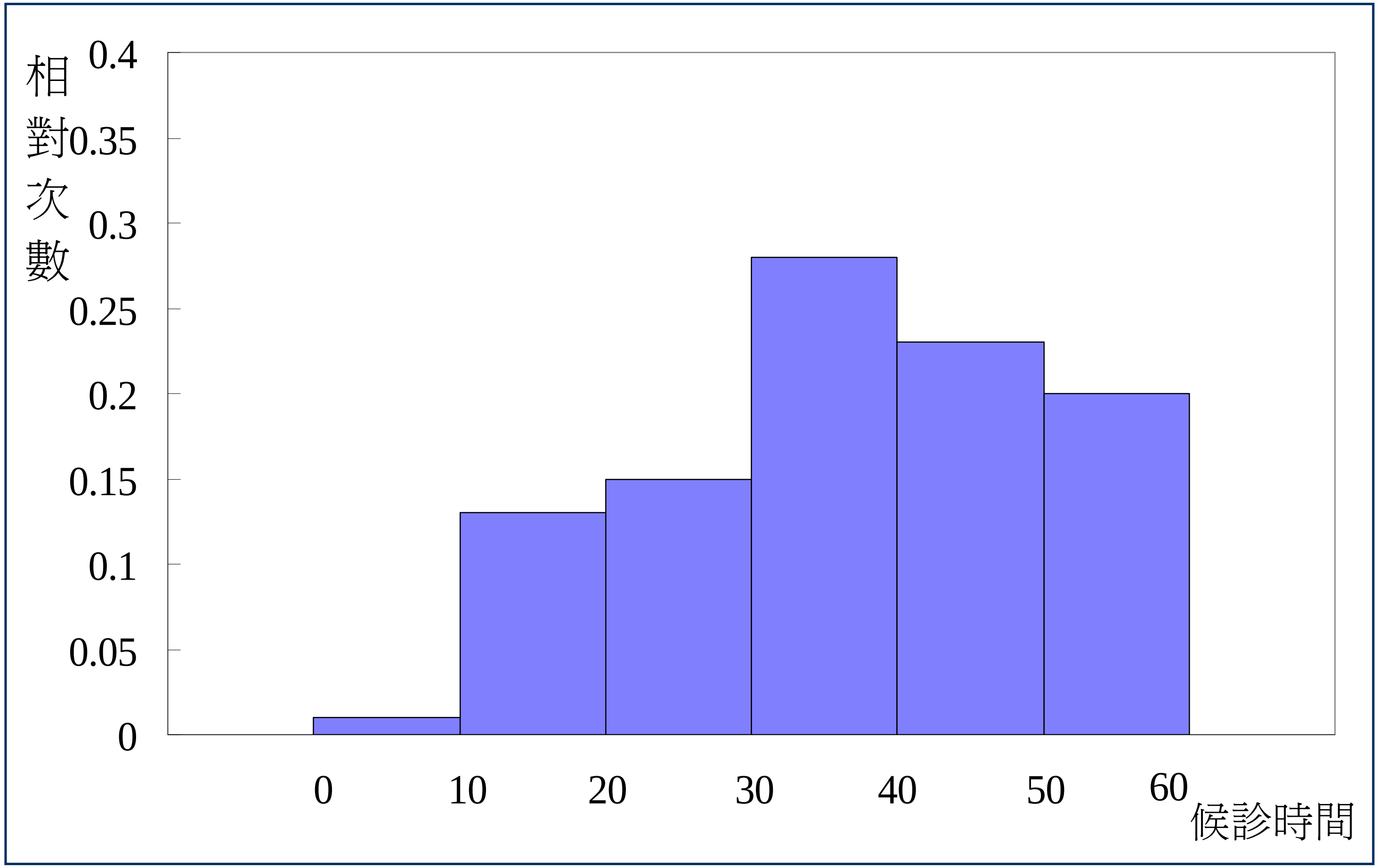
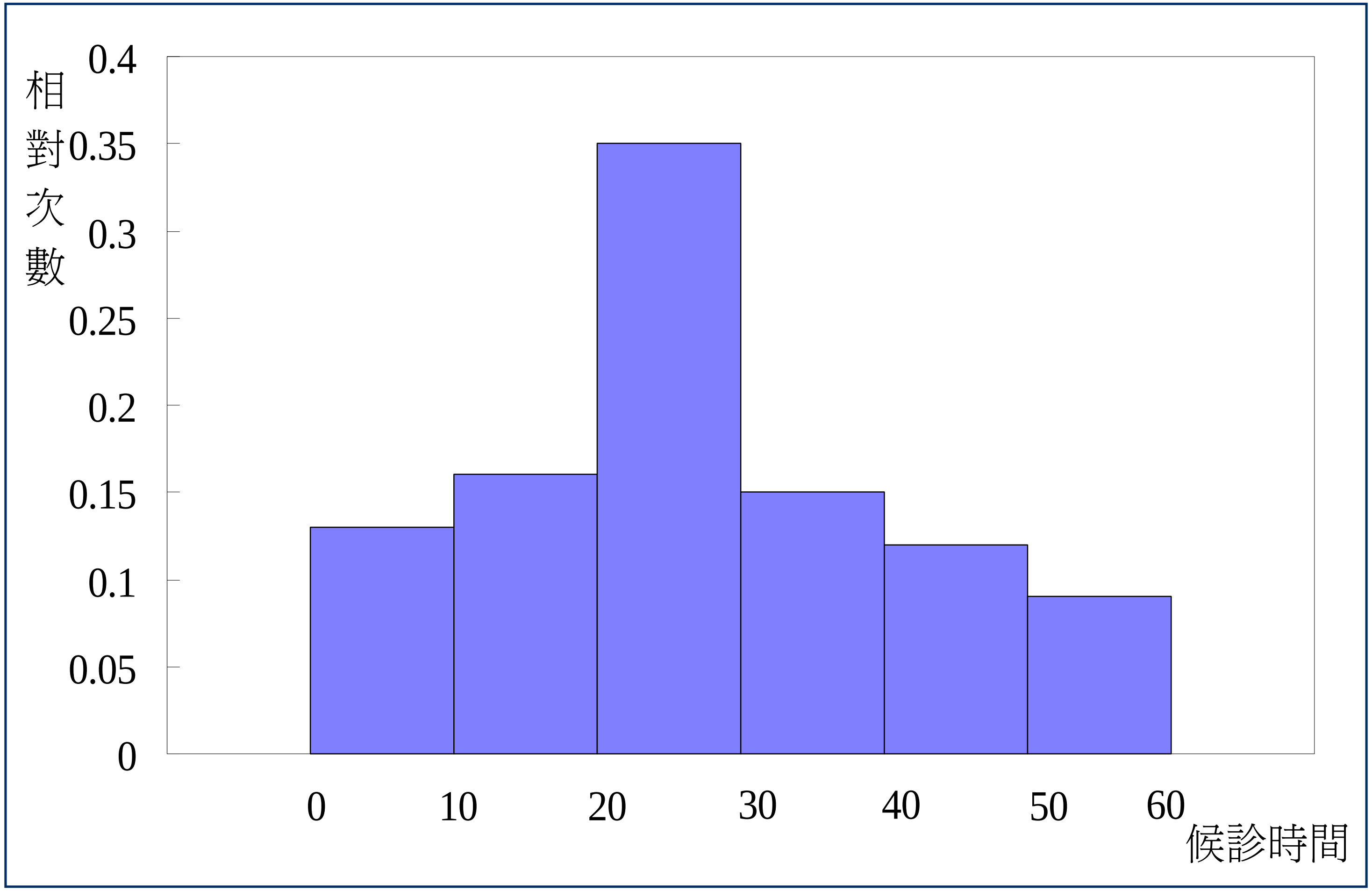


圖9.3 等待看牙時間(樣本2)

超過50分鐘的患者不到10%



9.1.2 抽樣誤差與非抽樣誤差

估計誤差(**error of estimation**): 無論抽樣發法如何科學, 樣本統計量與母體參數之間總是會有一些誤差

- **抽樣誤差 (sampling error)**

  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 (non-sampling error)**

  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 雙葉書廊發行 2004

- 抽樣誤差

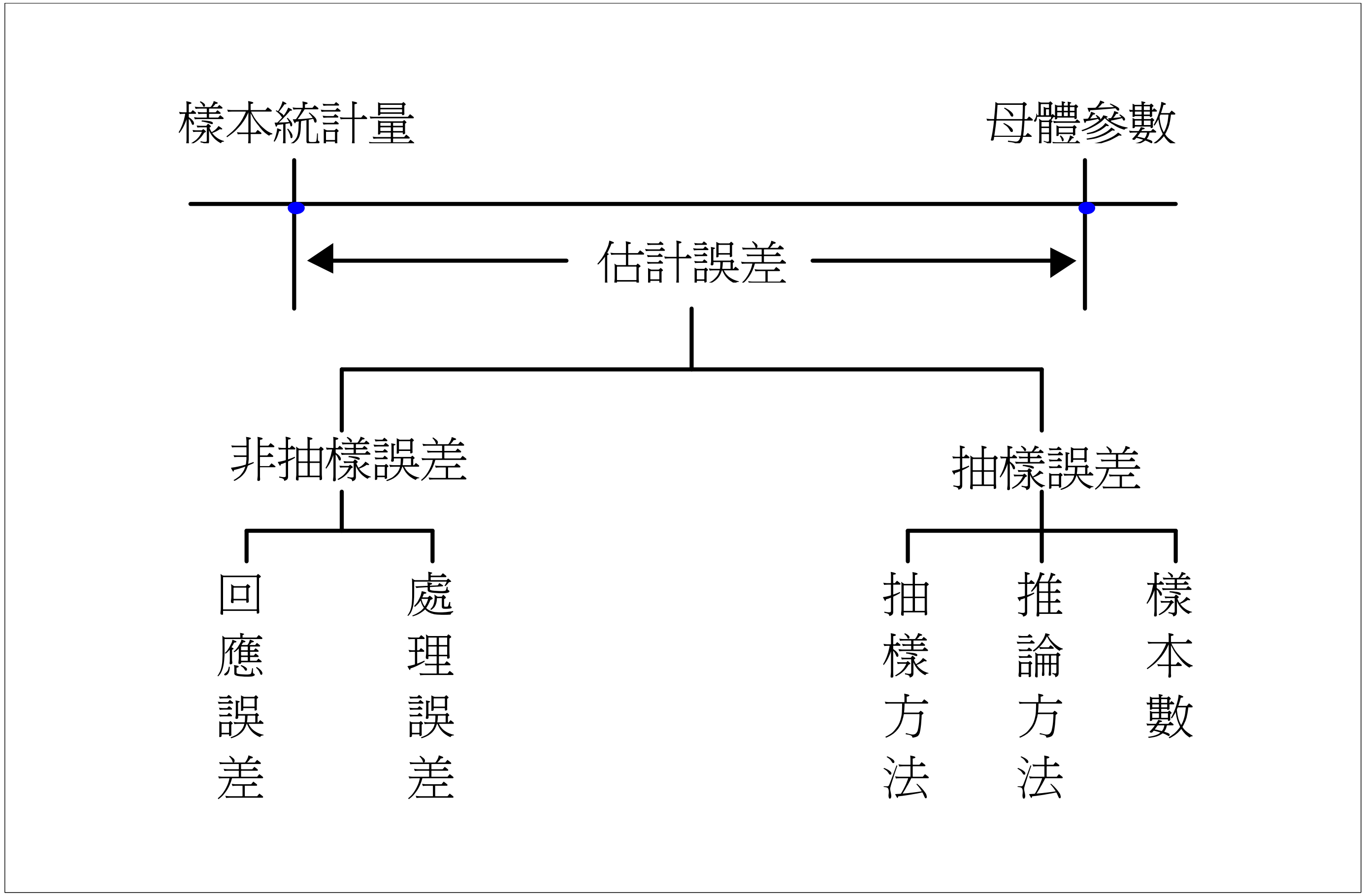
1. 選取樣本時, 機遇所造成, 即隨機抽樣誤差 (random sampling error)
   → 可經由樣本數的多寡來控制
2. 鎖使用的抽樣方法所造成
   → 機率抽樣法的抽樣誤差較小

※ 非抽樣誤差

1. 處理誤差 (processing error)
   處理資料時所犯的錯誤, 如計算錯誤, 輸入資料錯誤
   → 多方審核data
2. 回應誤差 (response error)
   受訪對象不願成實回答或對問題不了解,或不用心隨便回答

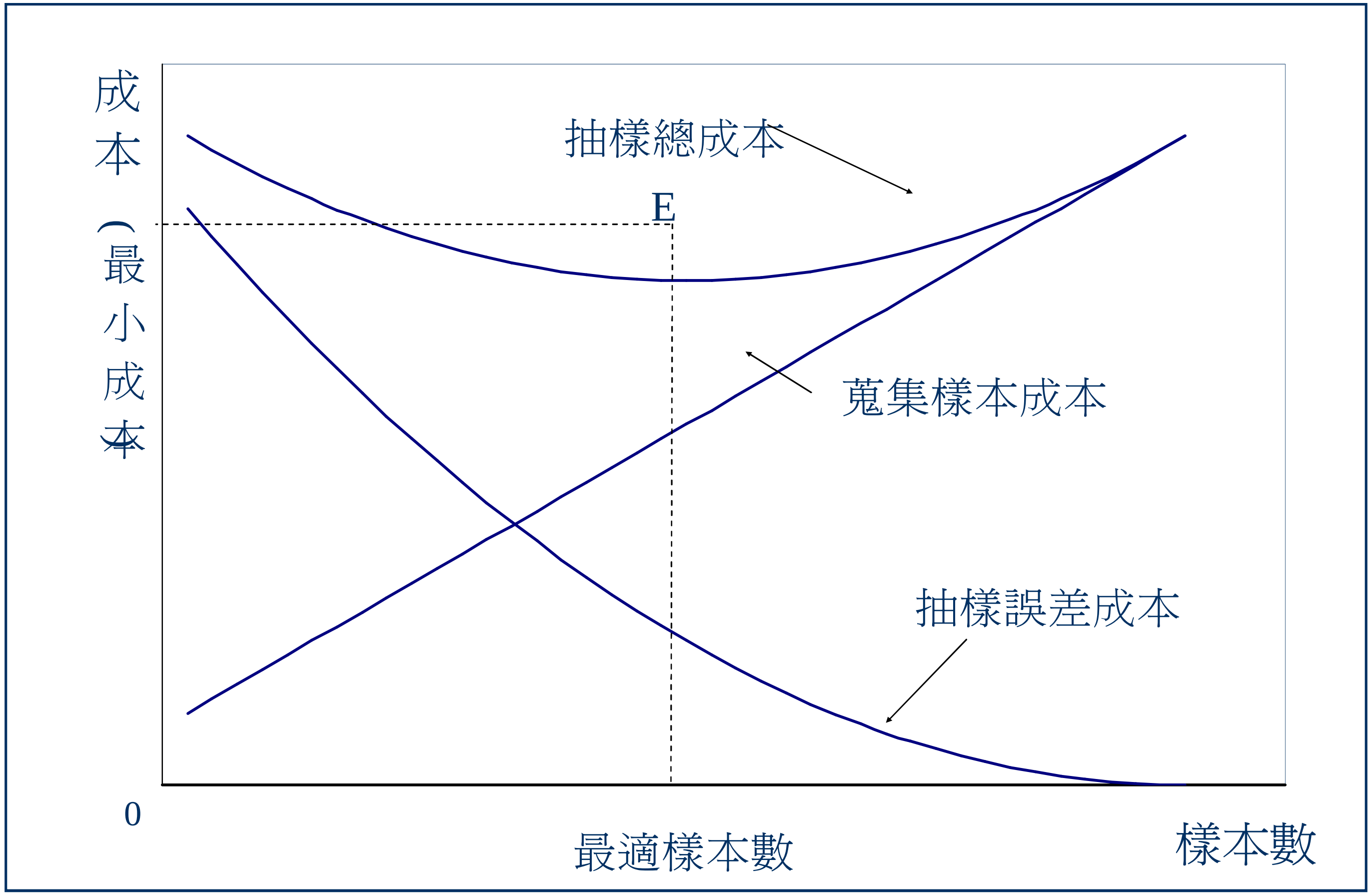
林惠玲　陳正倉著 雙葉書廊發行 2004

圖9.4 抽樣誤差與非抽樣誤差



9.1.3 抽樣成本與抽樣誤差

圖9.5 資料蒐集成本與抽樣誤差的關係



林惠玲 陳正倉著 雙葉書廊發行 2004

9.1.4 抽樣單位與抽樣底冊

抽樣單位 (sampling unit)

母體中的一個母體元素或一組母體元素

抽樣底冊 (sampling frame)

抽樣單位的名冊或一覽表

9.2 簡單隨機抽樣

9.2.1 簡單隨機抽樣的方法

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。

(1) 有限母體的簡單隨機抽樣

抽樣母體全部有N個元素(有限母體), 抽取n個樣本

a. 抽出不放回

b. 抽出放回

(2) 無限母體的的簡單隨機抽樣

隨機變數X, 其母體為無限, 機率分配為f(x)

林惠玲 陳正倉著 雙葉書廊發行 2004

9.2.2 簡單隨機抽樣的實施方式

- **簡單隨機抽樣的實施方式**

(1) 抽籤式

a. 抽出不放回，直到抽取n個為止

b. 抽出放回的方式，依序抽取n個

(2) 以亂數表抽取樣本

(3) 以電腦做隨機抽樣

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

圖9.6 數列對話方塊

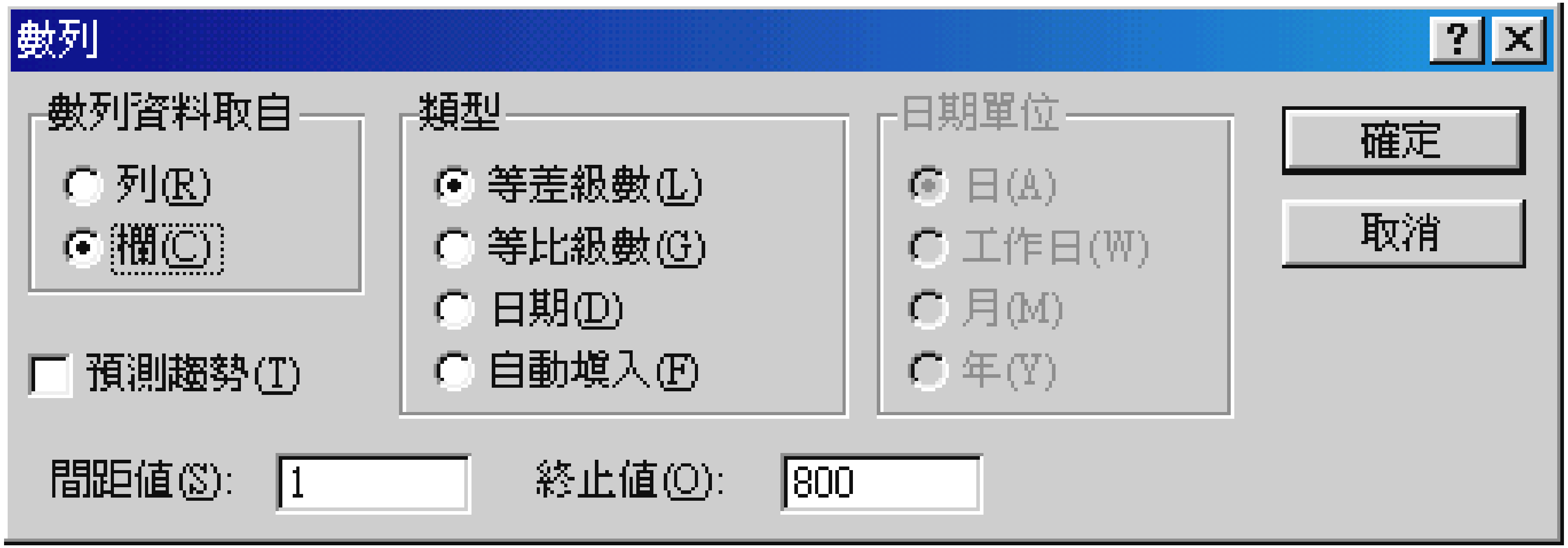
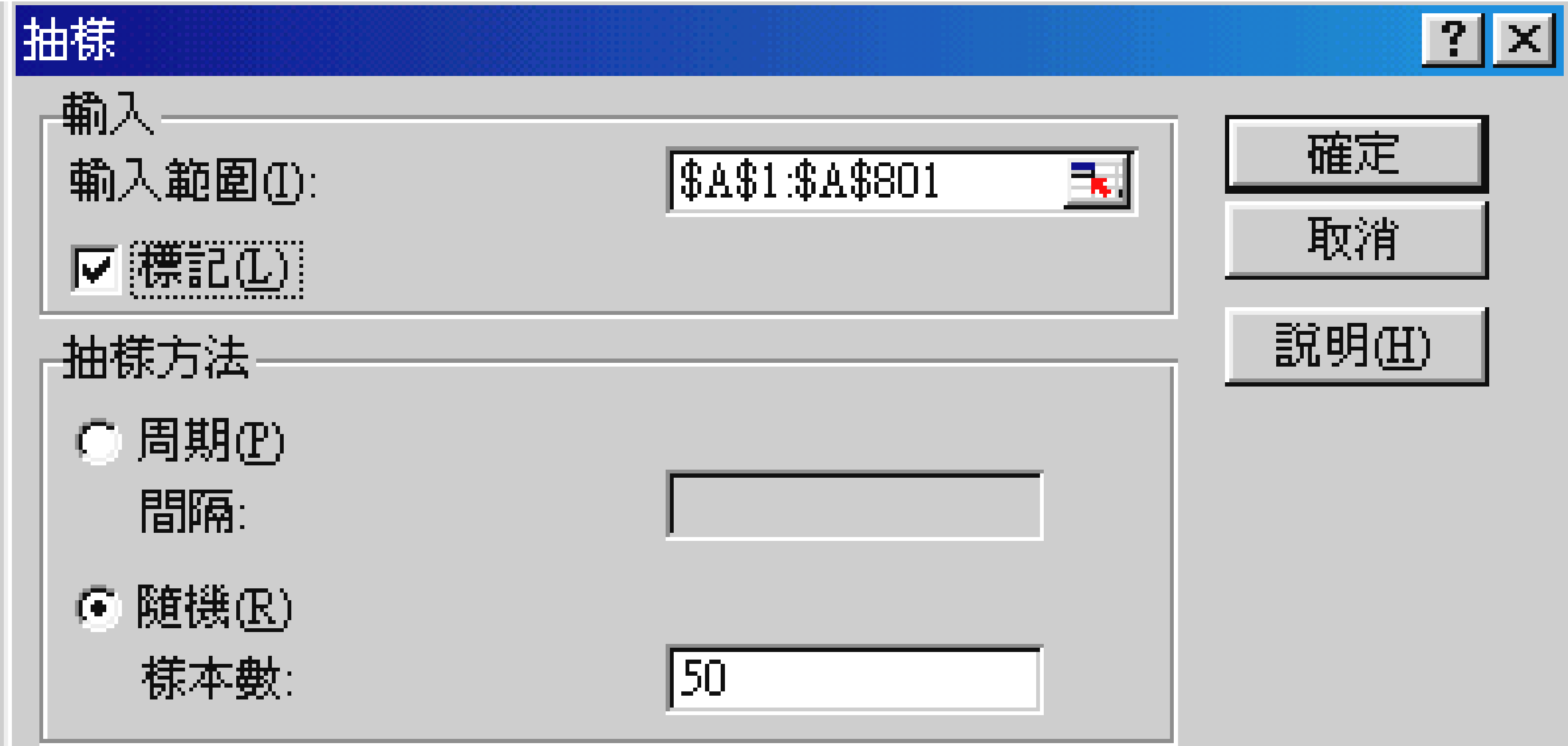

數列
數列資料取自
列(R)
欄(C)
預測趨勢(T)
類型
等差級數(L)
等比級數(G)
日期(D)
自動填入(F)
日期單位
日(A)
工作日(W)
月(M)
年(Y)
確定
取消
間距值(S):
1
終止值(O):
800

圖9.7 隨機抽樣對話方塊



9.3 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。

9.4 樣本平均數的抽樣分配

Concept

利用樣本統計量去推論母體參數值時, 使用的的樣本統計量是否能夠正確的估計母體參數?

→ 先了解統計量的值可能出現的情況

→ 統計量的機率分配

→ **抽樣分配 (sampling distribution)**

意義: 樣本統計量的機率分配稱為抽樣分配

9.4.1 母體分配 (population distribution)

母體分配是母體元素的機率分配

林惠玲 陳正倉著 雙葉書廊發行 2004

9.4.2 樣本平均數的抽樣分配

❍ 樣本平均數的抽樣分配

設母體爲隨機變數X，其機率分配爲 $f(x)$ ，若自母體中簡單隨機抽取n個元素爲一組樣本，表爲$(X_1, X_2, ..., X_n)$ ，若令

$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ ，則 $\overline{X}$ 爲樣本平均數。其機率分配表爲 $f(\overline{x})$ ，稱爲樣本平均數的抽樣分配。

林惠玲 陳正倉著 雙葉書廊發行 2004

假設中泰汽車公司有5個展示接待小姐, 月薪(千元)分別為 22 25 25 28 30

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
	$N = 5$

表9.3 展示小姐月薪的**母體**機率分配

x	$f(x)$
22	$1/5=0.2$
25	$2/5=0.4$
28	$1/5=0.2$
30	$1/5=0.2$
	$\sum f(x)=1$

林惠玲 陳正倉著 雙葉書廊發行 2004

母體平均數為

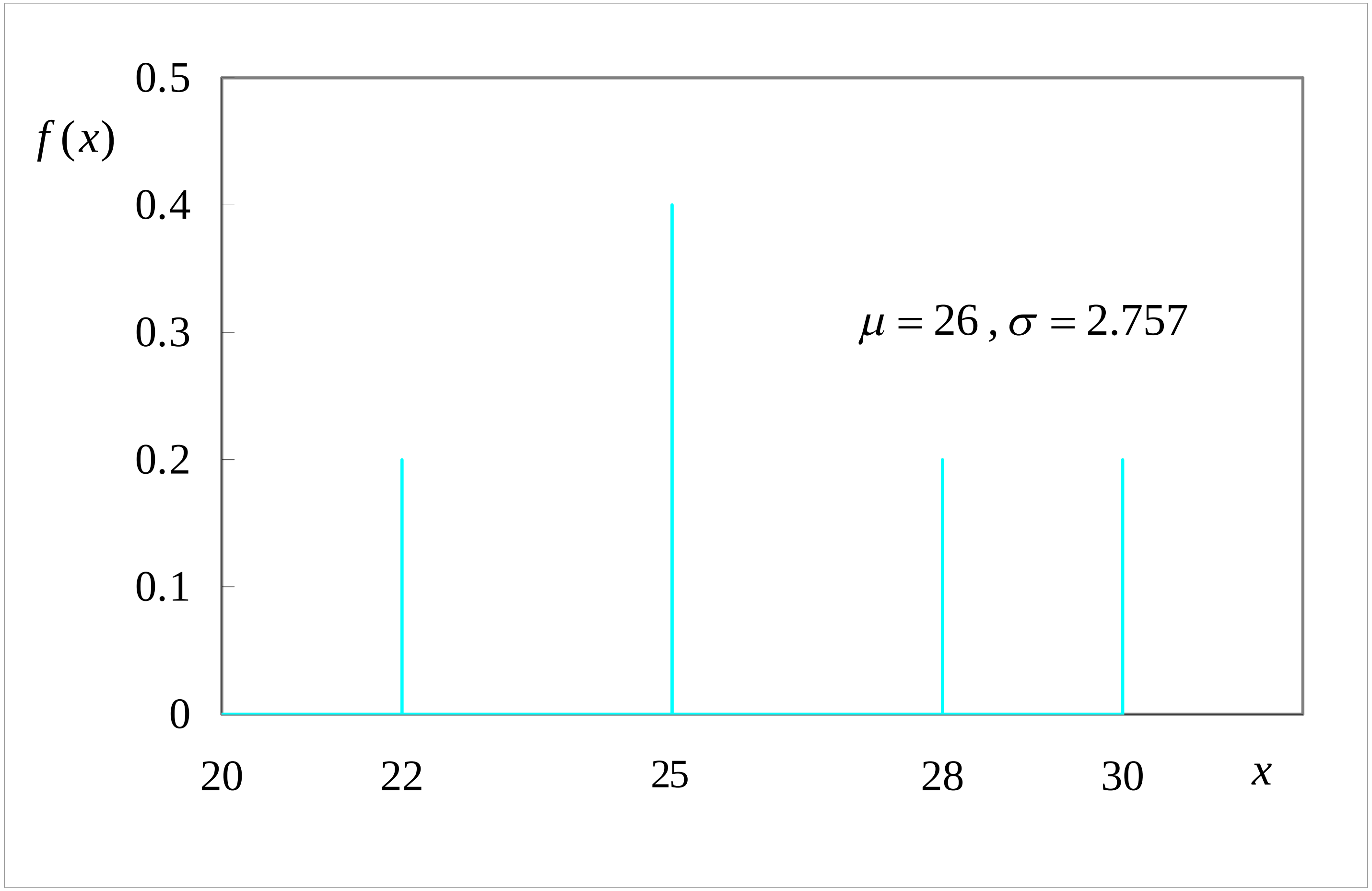
$$\mu = \frac{(22+25+25+28+30)}{5} = 26$$

變異數為

$$\sigma^2 = E(X^2) - [E(X)]^2 = 683.6 - 26^2 = 7.6$$

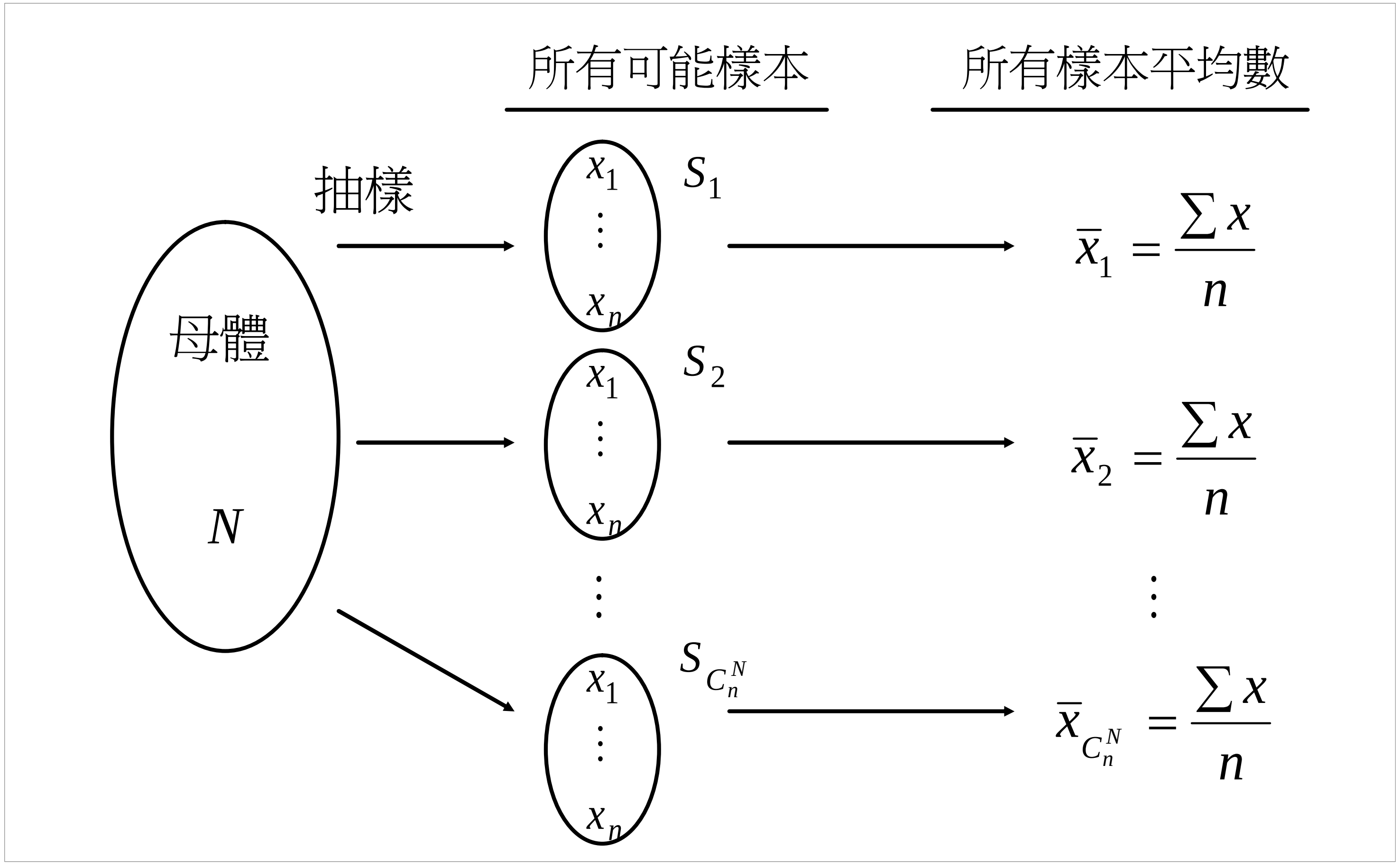
標準差為 σ=2.757

圖9.8 展示小姐月薪的母體機率分配



林惠玲 陳正倉著 雙葉書廊發行 2004

圖9.9 樣本平均數的抽樣分配



林惠玲 陳正倉著 雙葉書廊發行 2004

表9.4　樣本平均數的機率分配

$\bar{x}$	$f(\bar{x})$
$\bar{x}_1$	$1/C_n^N$
$\bar{x}_2$	$1/C_n^N$
$\vdots$	$\vdots$
$\bar{x}_{C_n^N}$	$1/C_n^N$
$\bar{X}$ 的平均數與變異數	$E(\bar{X})$、$V(\bar{X})$

抽出不放回

圖9.10　展示小姐月薪的抽樣

母　體		樣本空間		
A=22 B=25 C=25 D=28 E=30	→	(ABC) (ABD) (ABE) (ACD) (ACE) (ADE) (BCD) (BCE) (BDE) (CDE)	定義 →	$\overline{X}$:樣本組的平均數 $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$

林惠玲　陳正倉著 雙葉書廊發行 2004

表9.5 展示小姐月薪的樣本平均數

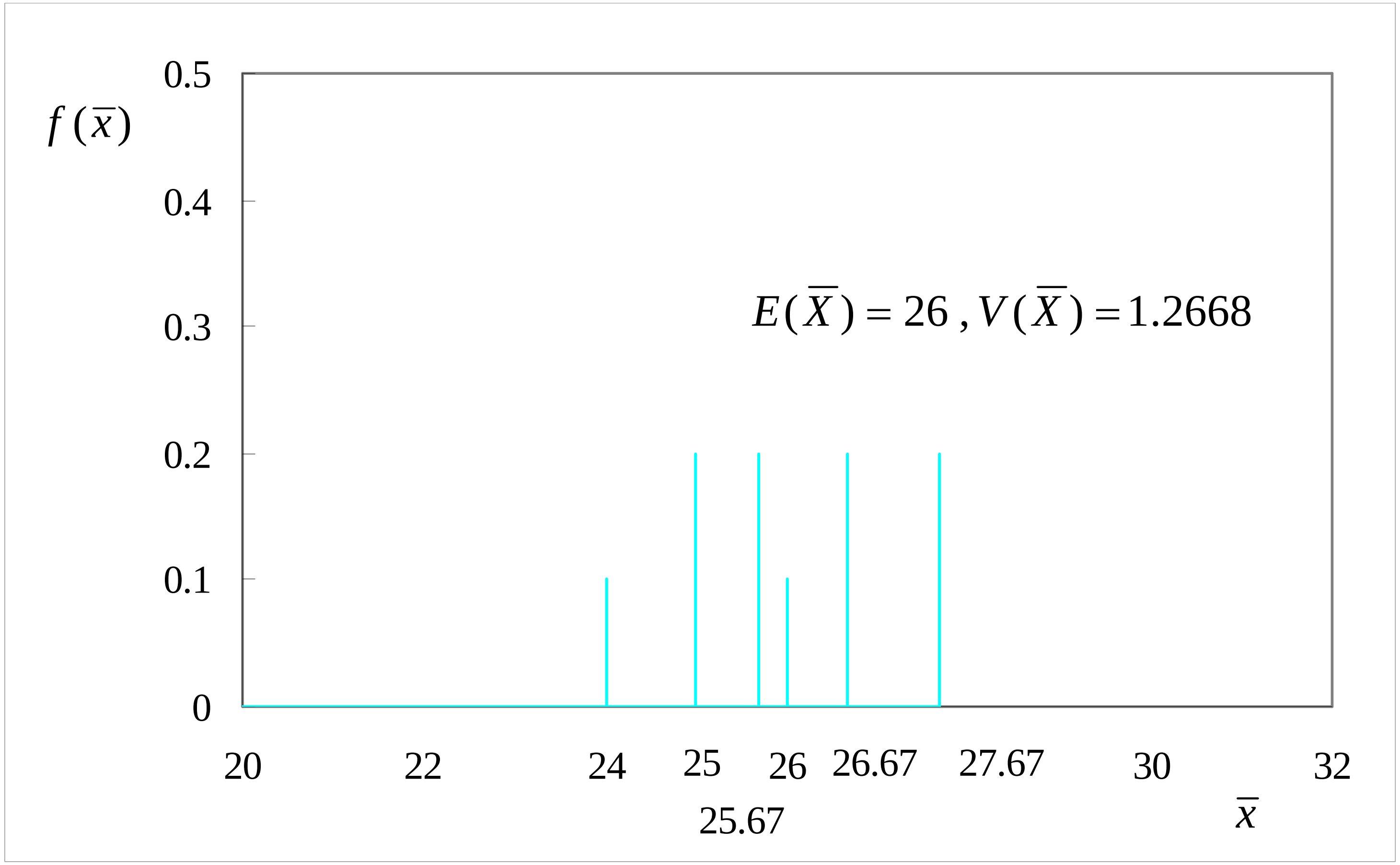
樣本	樣本平均數 $\bar{x}$
(ABC) = (22,25,25)	24
(ABD) = (22,25,28)	25
(ABE) = (22,25,30)	25.6667
(ACD) = (22,25,28)	25
(ACE) = (22,25,30)	25.66667
(ADE) = (22,28,30)	26.66667
(BCD) = (25,25,28)	26
(BCE) = (25,25,30)	26.66667
(BDE) = (25,28,30)	27.66667
(CDE) = (25,28,30)	27.66667

表9.6 展示小姐月薪的抽樣分配

$\bar{x}$	$\bar{x}^2$	$f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
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	$2/10 = 0.20$	142.2222578
27.66667	765.4446289	$2/10 = 0.20$	159.0999258
		$\sum f(\bar{x}) = 1.00$	677.2667733

林惠玲 陳正倉著 雙葉書廊發行 2004

圖9.11 展示小姐月薪的抽樣分配圖



9.4.3 抽樣誤差與非抽樣誤差

抽樣誤差 = $\left|\overline{X}-\mu\right|$

EX9.3 從中泰展示小姐薪資的母體中抽到樣本(ACD), 問該樣本組的平均數與母體平均數有否誤差? 如有, 誤差爲多少?

Solution:

$$\overline{X}=(22+25+28)/3=25$$

已知母體平均數 μ=26, 故抽樣誤差爲

抽樣誤差 = $\left|\overline{X}-\mu\right|=\left|25-26\right|=1$

假定樣本(ACD)中的25被誤記爲24, 其平均數爲

$$\overline{X}=(22+24+28)/3=24.66$$

誤差 = $\left|\overline{X}-\mu\right|=\left|24.66-26\right|=1.34$

抽樣誤差=1 千元

非抽樣誤差=0.34 千元

林惠玲 陳正倉著 雙葉書廊發行 2004

9.4.4 母體分配, 抽取方式與 $\overline{X}$ 的抽樣分配

Ex 9.4 6個秘書,年資分別為1,2,3,4,5,6年, 若抽2位來推估母體平均數, 高估或低估的機率為何? 估中的機率為何?

圖9.12 秘書小姐年資的機率分配圖

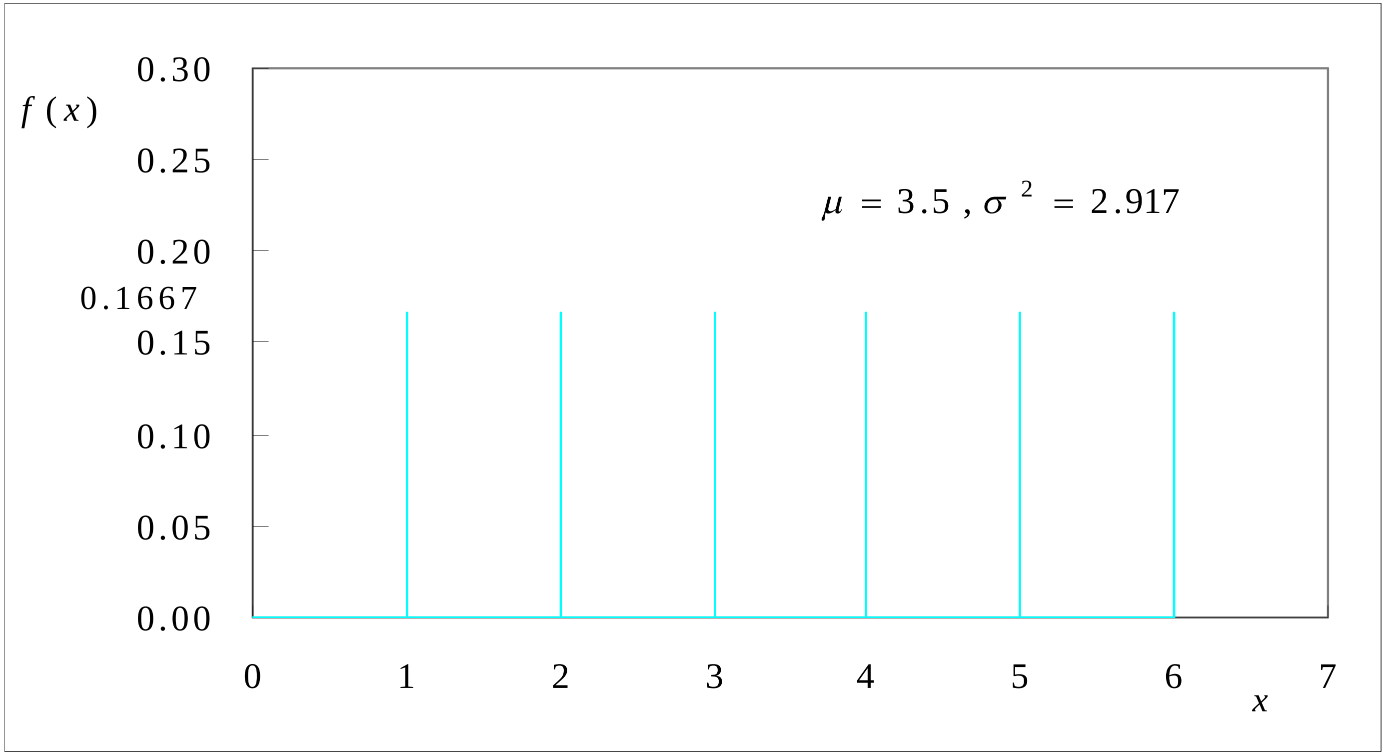


表9.7 秘書小姐年資的樣本平均數

樣本	樣本平均數	樣本	樣本平均數
(1,2)	1.5	(2,6)	4.0
(1,3)	2.0	(3,4)	3.5
(1,4)	2.5	(3,5)	4.0
(1,5)	3.0	(3,6)	4.5
(1,6)	3.5	(4,5)	4.5
(2,3)	2.5	(4,6)	5.0
(2,4)	3.0	(5,6)	5.5
(2,5)	3.5		

表9.8 秘書小姐的年資的抽樣分配

$\bar{x}$	$f(\bar{x})$
1.5	1/15
2.0	1/15
2.5	2/15
3.0	2/15
3.5	3/15
4.0	2/15
4.5	2/15
5.0	1/15
5.5	1/15

林惠玲 陳正倉著 雙葉書廊發行 2004

樣本平均數的平均數

$$E(\overline{X}) = \sum \overline{x} f(\overline{x}) = 3.5$$

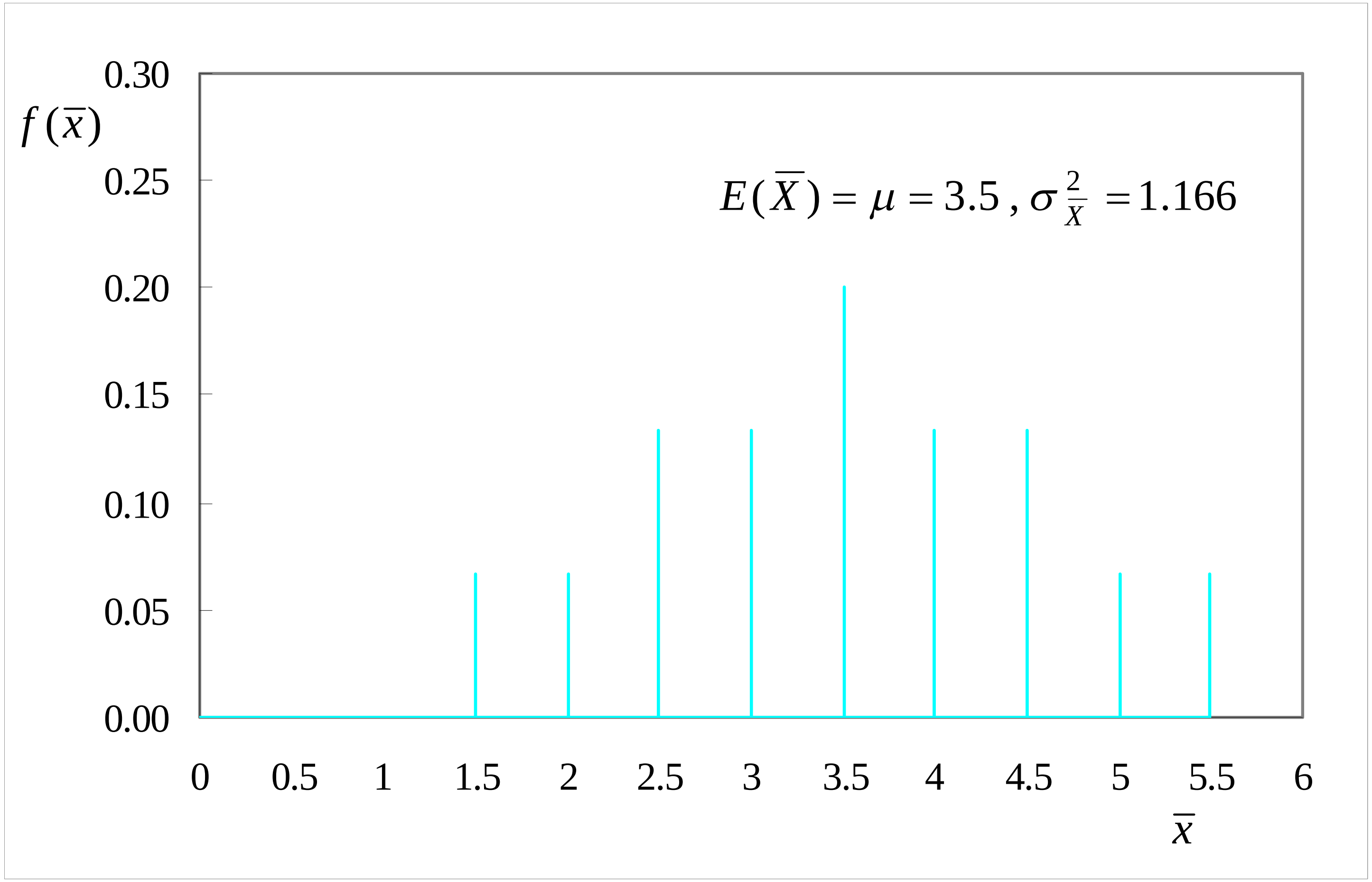
樣本平均數的變異數

$$V(\overline{X}) = \sum \overline{x}^2 f(\overline{x}) - (3.5)^2 = 13.416 - 12.25 = 1.166$$

見下圖

為一對稱分配, 樣本平均數等於母體平均數3.5年, 變異數等於1.166, 小於母體變異數2.917

圖9.13 秘書小姐的年資抽樣分配圖



EX9.5 設擲骰子2次 (等於抽取樣本數爲2的樣本), 令Xi表示第i 次骰子出現的點數 , i=1,2. 則樣本平均數爲 $\bar{X} = \frac{X_1 + X_2}{2}$
問 $\bar{X}$ 的機率分配, 平均數與變異數爲何?

Solution:

此時母體爲一均等分配

f(x)=1/6, x=1,2,…,6

並知

母體平均數爲 $\mu = E(X) = \sum xf(x) = 3.5$

母體變異數爲

$$\sigma^2 = V(X) = \sum (x-\mu)^2 f(x) = 2.917$$

圖9.14 擲骰子出現點數的機率分配圖

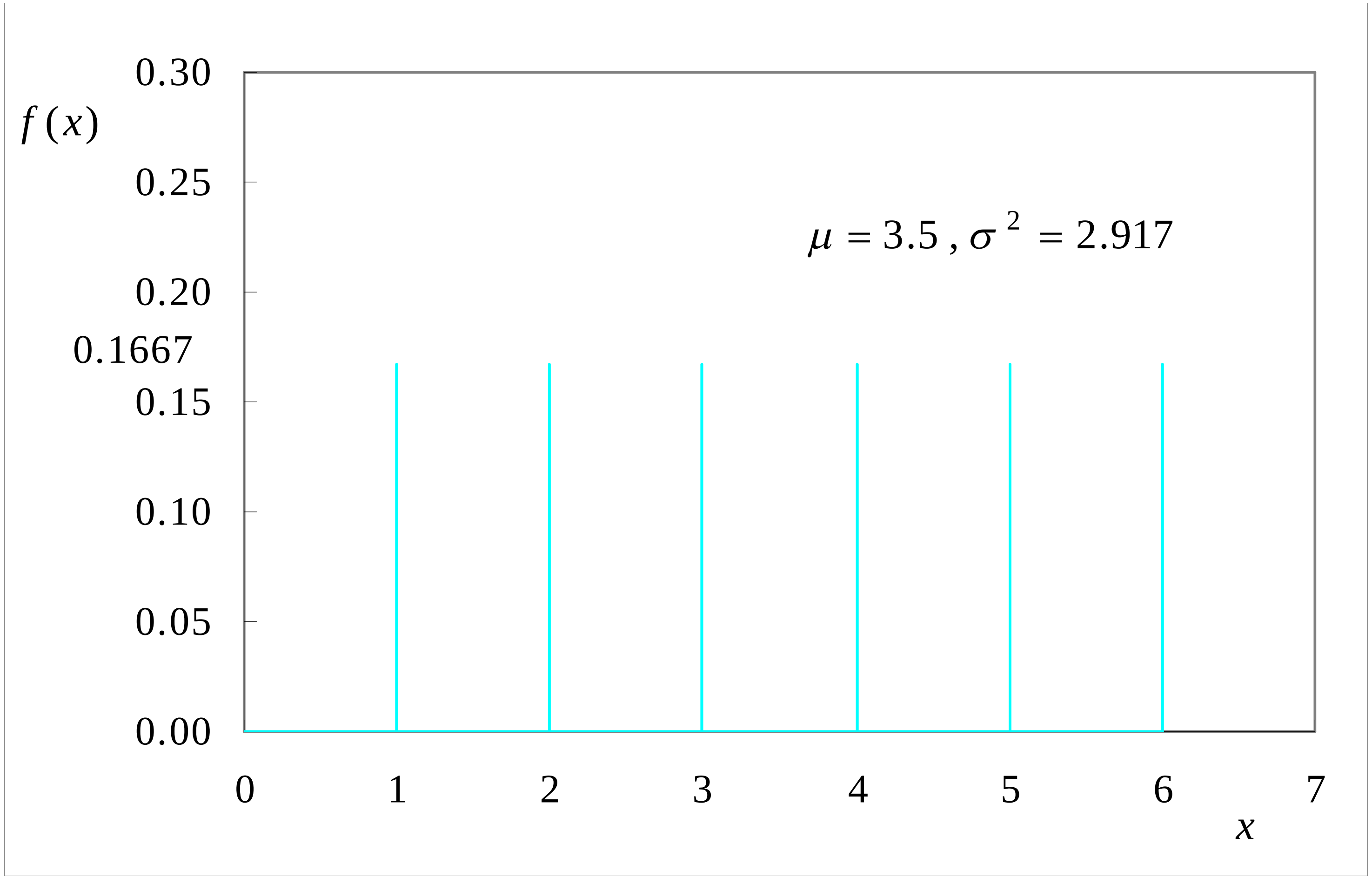


表9.9　擲骰子兩次的樣本平均數的抽樣分配

樣本	$\bar{x}$	$f(\bar{x})$
(1，1)	1	1/36
(1，2)(2，1)	3/2	2/36
(1，3)(3，1)(2，2)	4/2	3/36
(1，4)(4，1)(2，3)(3，2)	5/2	4/36
(1，5)(5，1)(2，4)(4，2) (3，3)	6/2	5/36
(1，6)(6，1)(2，5)(5，2)(3，4)(4，3)	7/2	6/36
(2，6)(6，2)(3，5)(5，3)(4，4)	8/2	5/36
(3，6)(6，3)(4，5)(5，4)	9/2	4/36
(4，6)(6，4)(5，5)	10/2	3/36
(5，6)(6，5)	11/2	2/36
(6，6)	12/2	1/36

圖9.15 擲骰子兩次樣本平均數的抽樣分配圖

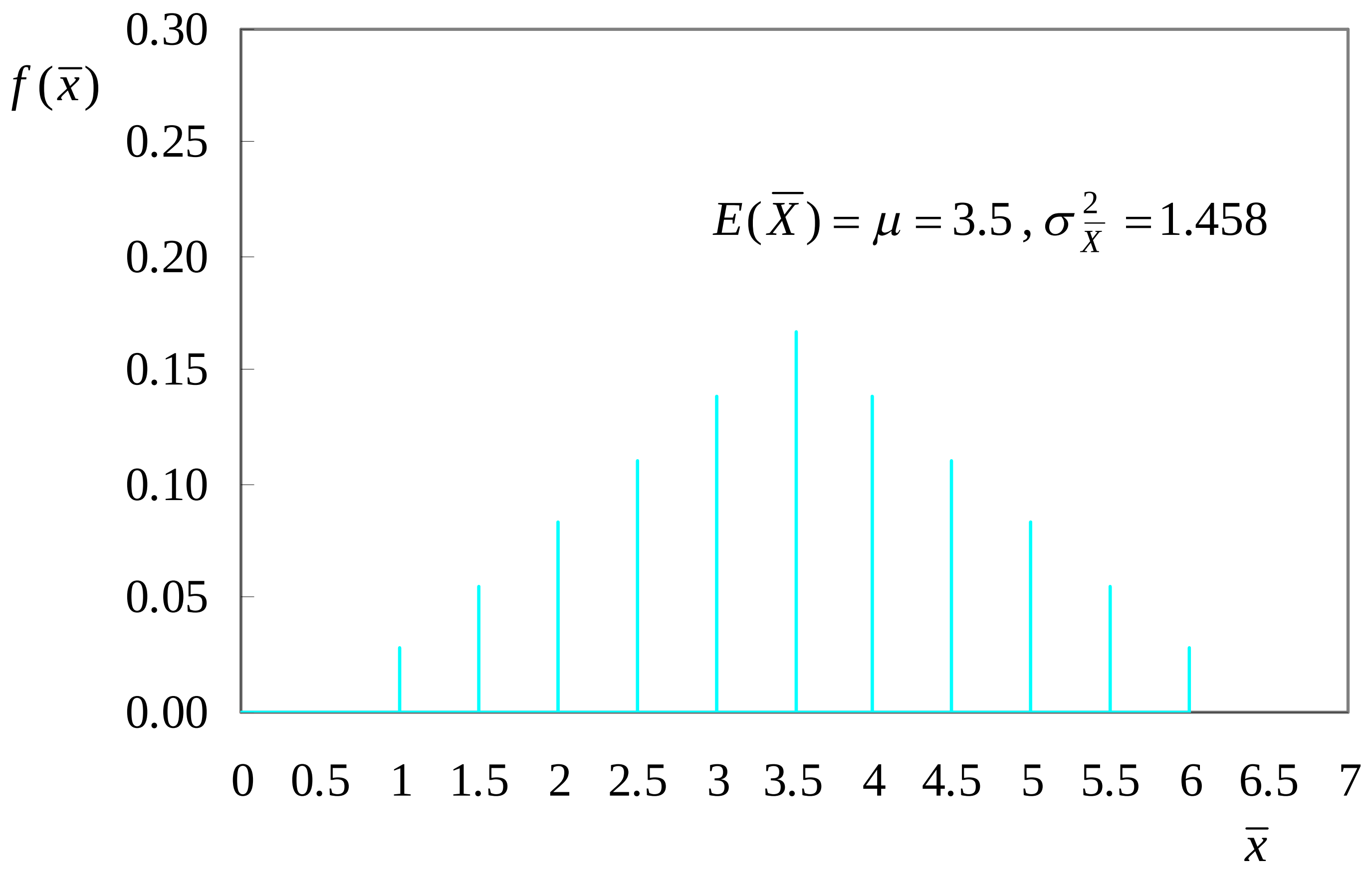


表9.10 三個抽樣分配的比較

母體分配	母體平均數	母體變異數	抽取方式	$\overline{X}$ 平均數	$\overline{X}$ 變異數	$\overline{X}$ 的分配
偏態分配	26	7.6	抽出不放回	26	1.373	雙峰分配
均等分配	3.5	2.917	抽出不放回	3.5	1.166	對稱分配
均等分配	3.5	2.917	抽出放回	3.5	1.458	對稱分配

9.4.5 樣本平均數的平均數與變異數

- **$\overline{X}$ 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爲 $\overline{X}$ 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($\sigma_{\overline{X}}$)**

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

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

9.4.6 樣本平均數的平均數與變異數的特性

1. $\bar{X}$ 的抽樣分配以母體平均數μ為中心, 無論樣本數大小, 母體有限或無限, 抽出放回或不放回, $\bar{X}$ 的抽樣分配的平均數為μ.

2. $\bar{X}$ 的抽樣分配的變異數受到母體變異數 (σ^2), 樣本數(n)以及母體為有限或無限的影響

   a. 當母體變異數越大, $\bar{X}$ 的抽樣分配的變異數亦大

   b. 當樣本數n增加時, $V(\bar{X})$ 愈小, 抽樣分配愈集中於母體平均數μ .

→ $n \to \infty,\ V(\bar{X}) = \dfrac{\sigma^2}{n} \to 0$ $\bar{X}$ 的機率分配收斂到母體平均數μ

林惠玲 陳正倉著 雙葉書廊發行 2004

大數法則 (law of large numbers)

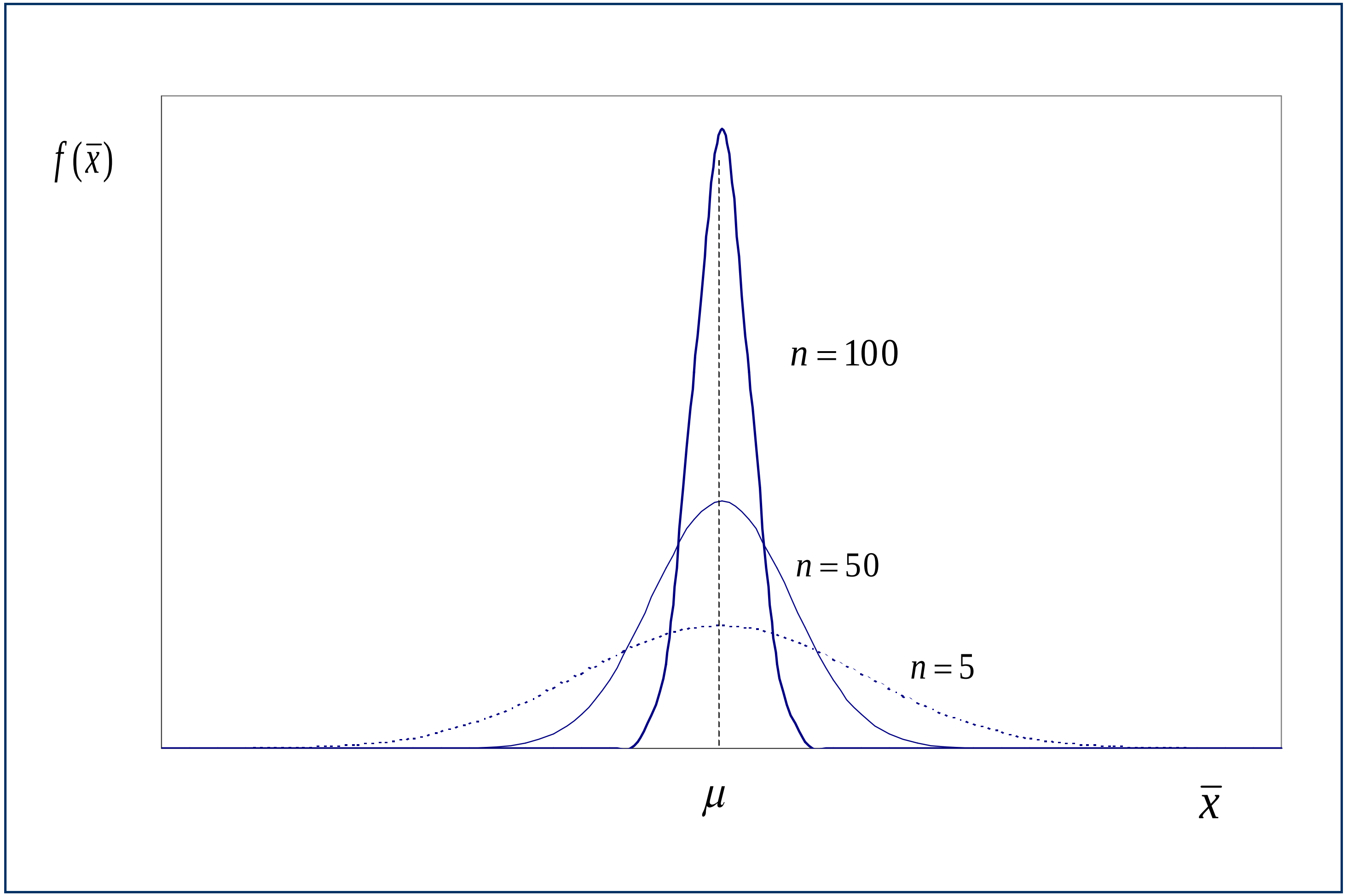
- **大數法則**

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

亦即當樣本數夠大時，$\bar{X}$ 會趨近於 μ 的機率為1。

林惠玲 陳正倉著 雙葉書廊發行 2004

圖9.16　大數法則

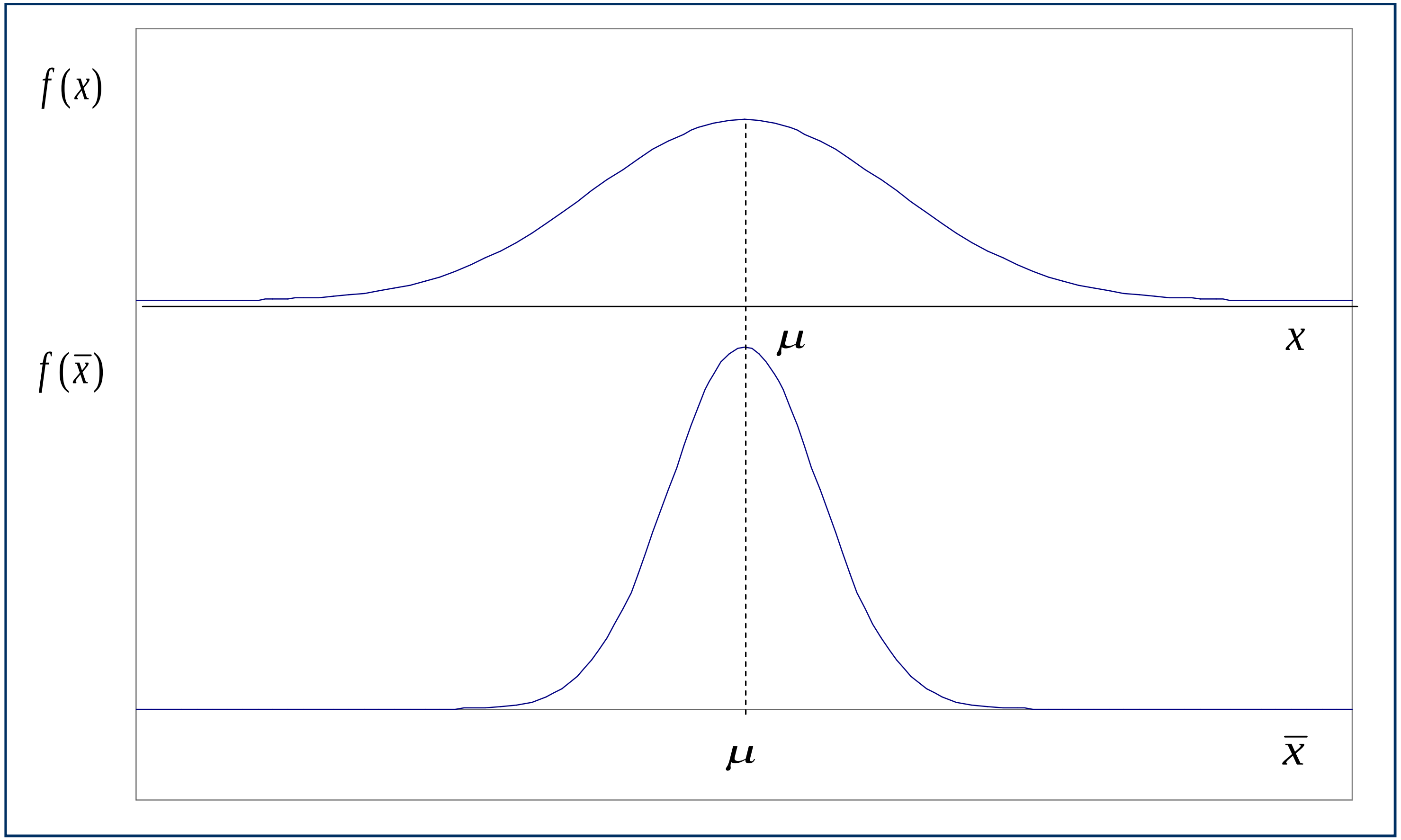


c. 母體爲有限, 且樣本數相對於母體元素個數不是很小($n/N > 0.05$), 則 $\overline{X}$ 的抽樣分配的變異數必須乘上有限母體修正因子 $(N-n)/(N-1)$

d. 有限母體的情況下, 若且樣本數相對於母體元素個數很小($n/N \le 0.05$), 或抽樣時採抽出放回的方式, $\overline{X}$ 抽樣分配的變異數與無限母體相同

3. 母體隨機變數X的機率分配比 $\overline{X}$ 的抽樣分配分散度大見下圖

圖9.17　X 的機率分配與 $\overline{X}$ 的抽樣分配



EX9.7 設X代表國中生的身高, 其母體平均數爲160公分, 標準差爲10公分, 現自所有國中生隨機抽取10個人, 則該10個人的平均身高(樣本平均數)的平均數與變異數爲何?

Solution:

樣本平均數與變異數如下

$$E(\overline{X}) = \mu = 160$$

$$V(\overline{X}) = \frac{\sigma^2}{n} = \frac{100}{10} = 10$$

EX 9.8 注意有限母體

9.4.7 抽樣分配與樣本大小的關係

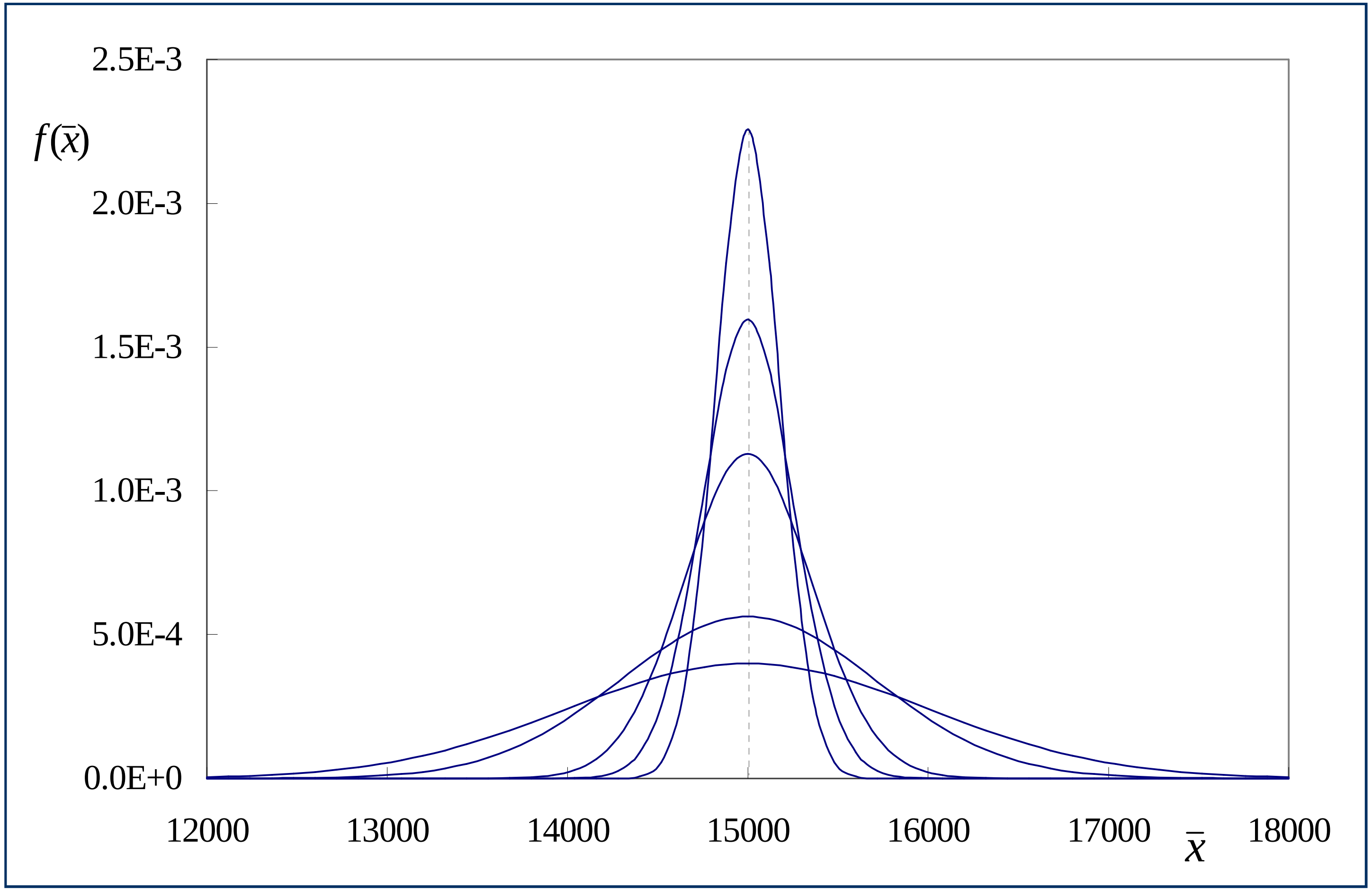
已知電源開關使用次數的平均數為15000次,標準差為1000, 現抽取樣本為 n=1,2,8,16,32, 樣本平均數的平均數與標準差為何?

表9.11　樣本數不同的抽樣分配

樣本數	$E(\overline{X}) = \mu$	$\sigma_{\overline{X}} = \sigma / \sqrt{n}$
$n = 1$	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	$1000 / \sqrt{32} = 176.78$

林惠玲　陳正倉著 雙葉書廊發行 2004

圖9.18 樣本大小不同的抽樣分配



林惠玲 陳正倉著 雙葉書廊發行 2004

9.4.8 常態母體樣本平均數的抽樣分配

常態母體 $\overline{X}$ 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 $\overline{X}$ 的抽樣分配亦為常態分配，其平均數和標準差分別為：

$$\mu_{\overline{X}} = \mu \text{ , } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad \overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

林惠玲 陳正倉著 雙葉書廊發行 2004

EX9.9 高雄市計程車通常每天耗用的燃料費爲一常態分配, 平均數爲313元, 標準差爲101元

1. 在高雄市隨機抽取一輛計程車,其每天耗用的燃料費超過400元的機率爲多少?
2. 某家計程車行有8輛計程車,則8輛計程車每天平均的燃料費爲多少元? 其鵬均值超過400元的機率爲何?

Solution

$X \sim N(313, 101^2)$

1.

$$P(X > 400) = P(Z > \frac{400-313}{101}) = P(Z > 0.86) = 0.1949$$

2.

$$\bar{X} \sim N(313, \frac{101^2}{8})$$

$$P(\bar{X} > 400) = P(Z > \frac{400-313}{101/\sqrt{8}}) = P(Z > 2.44) = 0.0119$$

林惠玲 陳正倉著 雙葉書廊發行 2004

9.5 中央極限定理 (Central limit theorem)

- **中央極限定理(非常態母體的抽樣分配)**

無論母體爲何種分配，其平均數爲μ, 變異數爲 σ^2

自母體簡單隨機抽取*n*個爲一組樣本，若樣本數*n*夠大(一般認爲)，則樣本平均數的抽樣分配會趨近於常態分配, 即

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

林惠玲 陳正倉著 雙葉書廊發行 2004

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，不論母體爲何種分配，當 $n \geq 30$ 時，$\overline{X}$ 漸趨於常態分配。

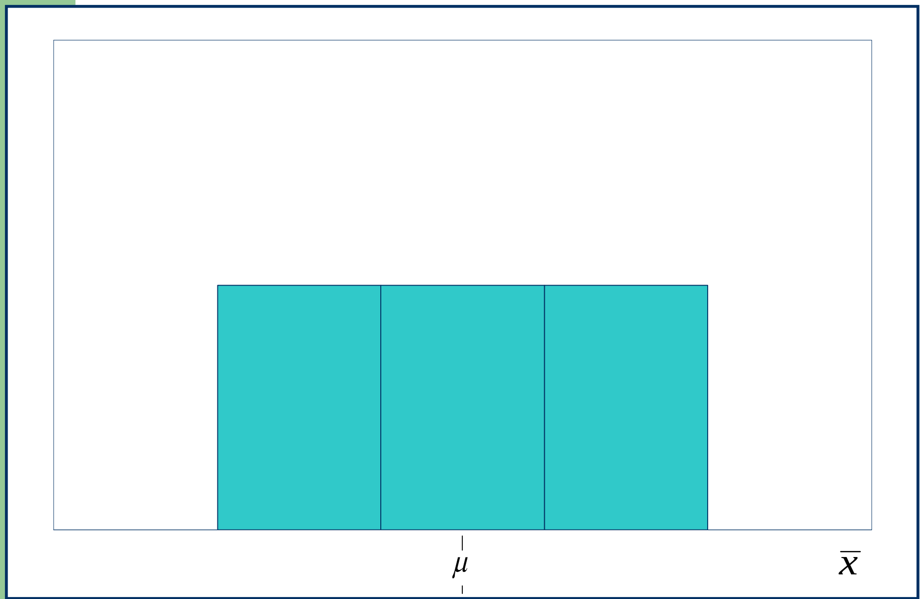
② 中央極限定理可適用於樣本統計量爲隨機樣本 $(X_1, X_2, \cdots, X_n)$ 線性函數的情況。

③若母體爲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。

圖9.19 中央極限定理

母體分配



母體分配

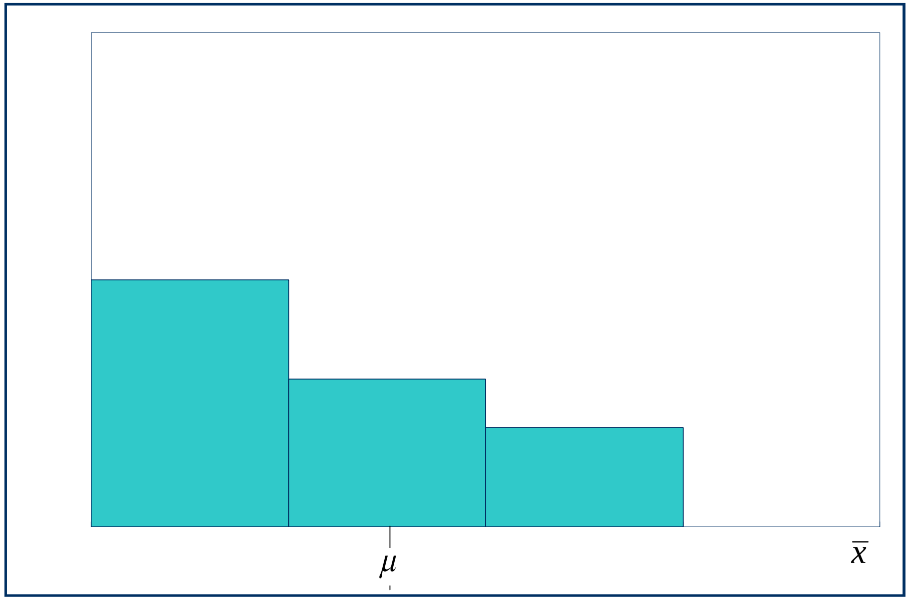


圖9.19 中央極限定理（續）

抽樣分配　　抽樣分配

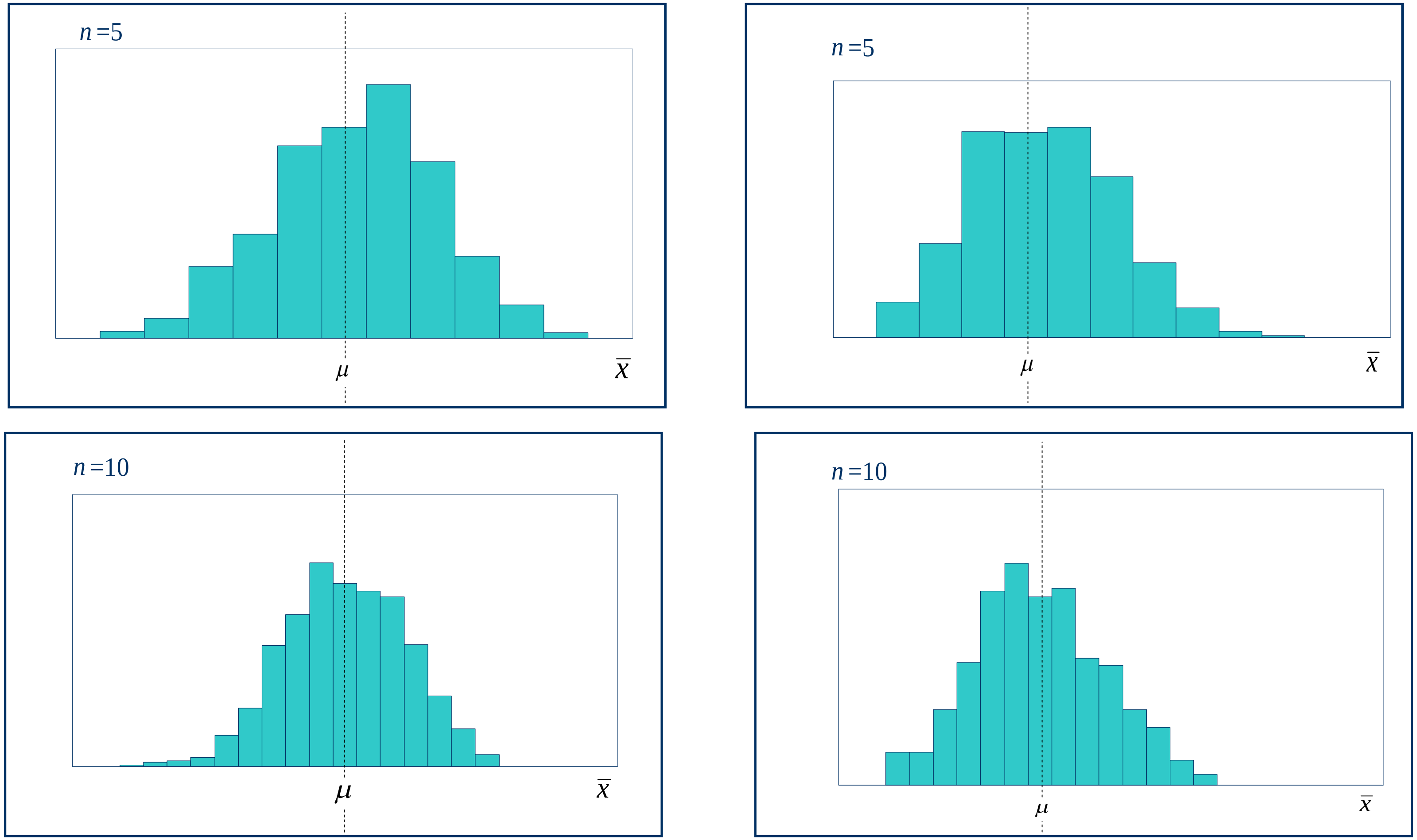


圖9.19　中央極限定理（續）

抽樣分配　　抽樣分配

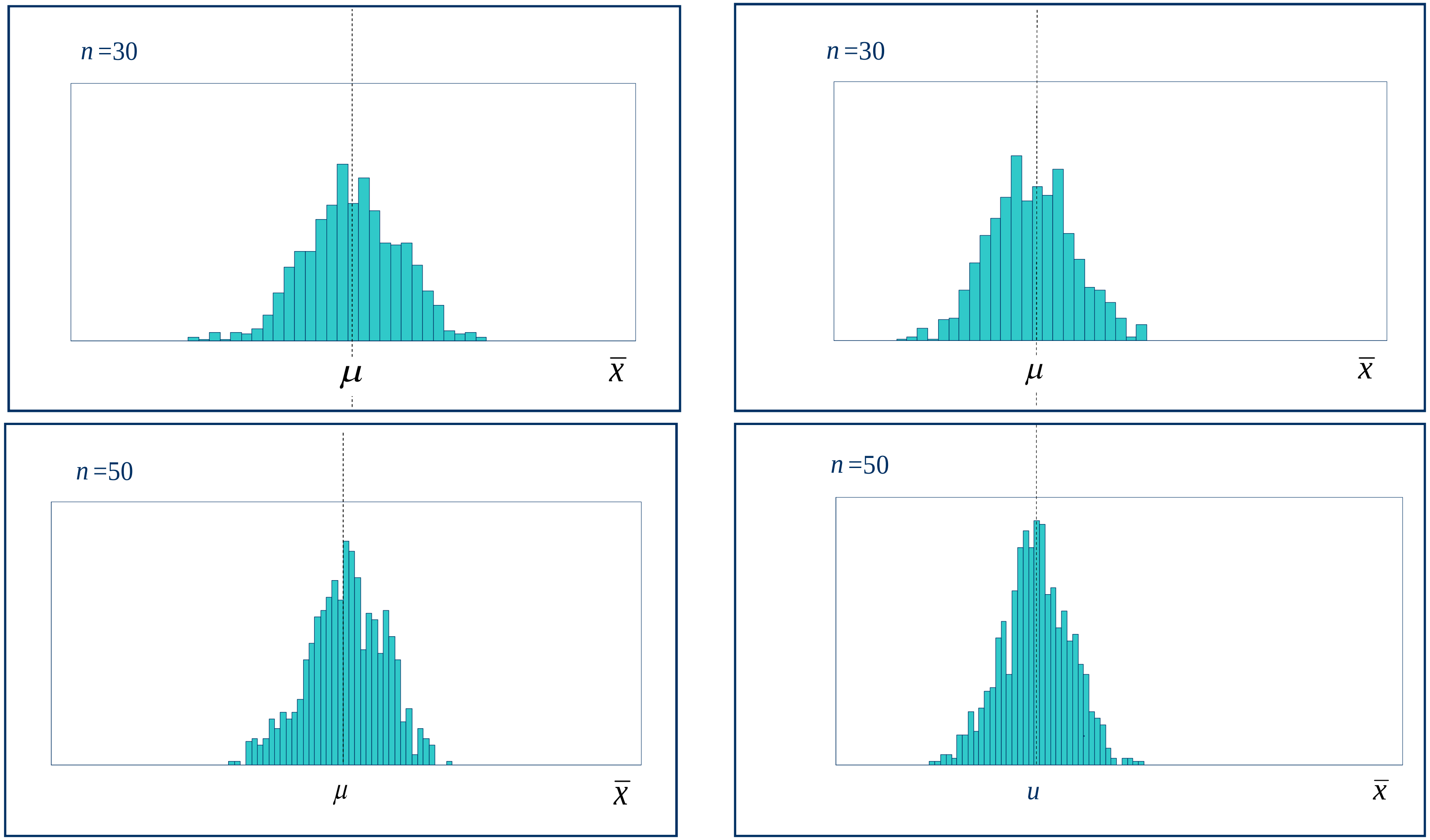


表9.12　$\overline{X}$的抽樣分配

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

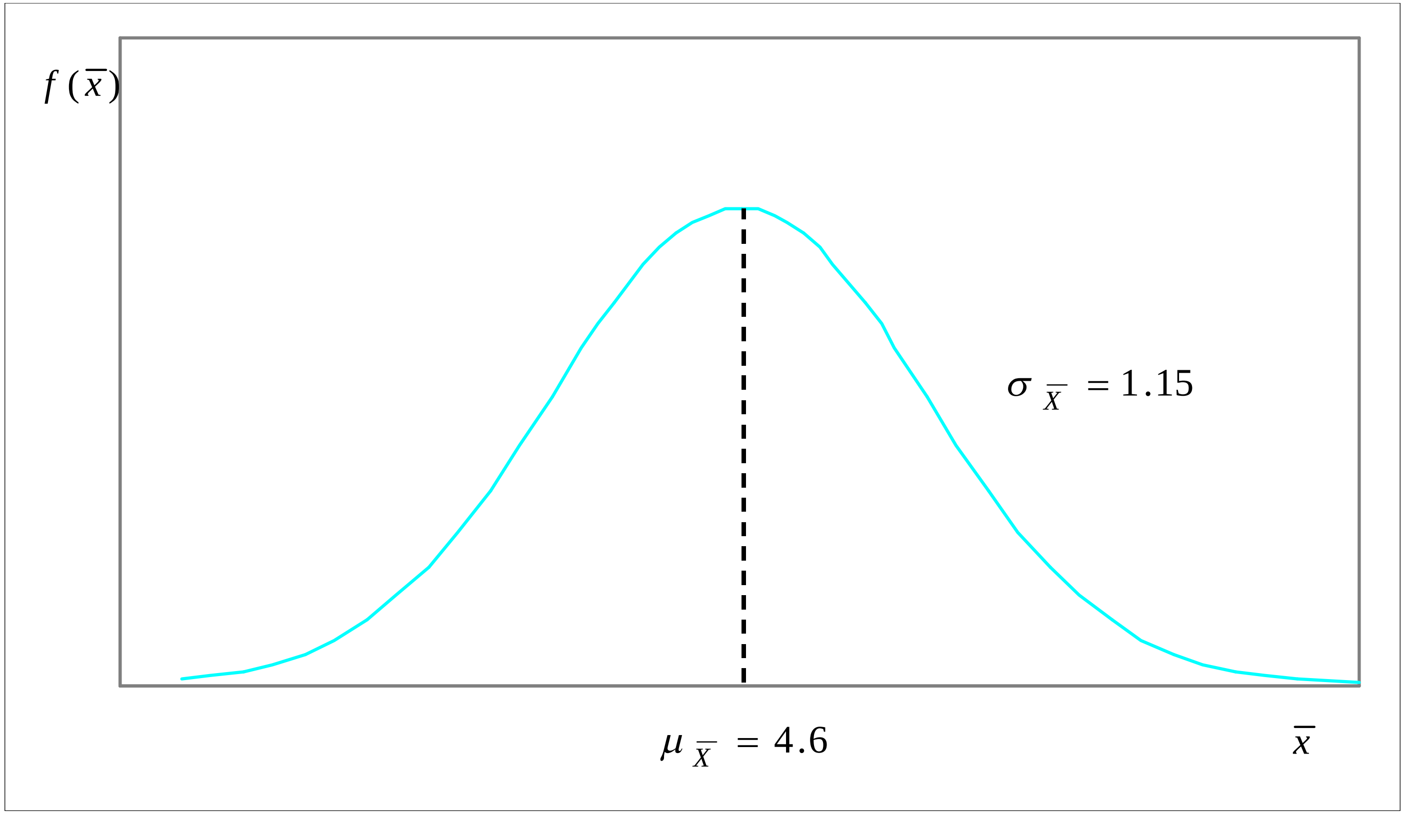
註：①若母體爲有限母體，且 $\frac{n}{N} > 0.05$，則 $V(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$。

②若母體爲有限母體，且 $n/N > 0.05$，則 $\overline{X}$ 不一定爲常態分配，因 $(X_1, \cdots, X_n)$ 不獨立。

林惠玲　陳正倉著 雙葉書廊發行 2004

EX 9.10 民國89年有5,179家紡織廠, 銷售額平均4.6百萬, 標準差11.5百萬

圖9.20 紡織業平均銷售額的抽樣分配



林惠玲 陳正倉著 雙葉書廊發行 2004

Q: 隨機抽取50及100家, 問 $\overline{X}$ 的平均數與標準差及抽樣分配的形狀爲何?

Solution:

1. 50家

$$\mu_{\overline{X}} = \mu = 4.6$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{11.5}{\sqrt{50}} = 1.63$$

2. 100家

$$\mu_{\overline{X}} = \mu = 4.6$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{11.5}{\sqrt{100}} = 1.15$$

- 樣本和的抽樣分配

令S爲隨機樣本的和, $S = X_1 + X_2 + \ldots + X_n$

樣本和的平均數爲
$$E(S) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = n\mu$$

樣本和的變異數

1. 若爲無限母體,或有限母體抽出放回

$$V(S) = V(X_1 + X_2 + \ldots + X_n) = n\sigma^2$$

2. 若爲有限母體抽出不放回

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

林惠玲 陳正倉著 雙葉書廊發行 2004

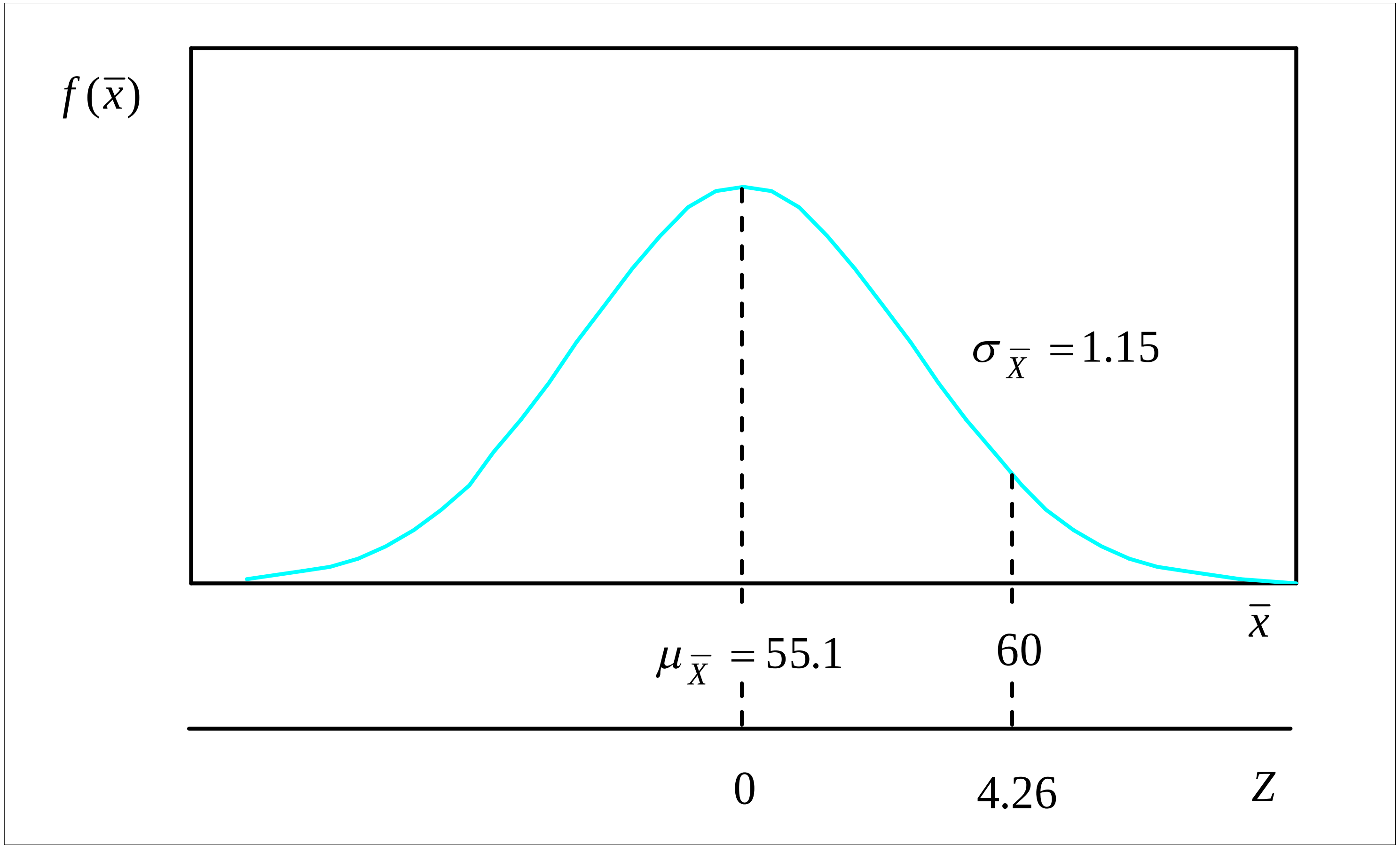
9.6 樣本平均數抽樣分配的應用

EX9.14 根據調查資料得知, 搭乘台鐵的乘客數雖逐年增加, 但每人平均乘車公里數卻反而逐年下降. 91年每人平均乘車公里數爲55.1公里, 較72年最高峰時的65.5公里減少10.4公里. 若每人平均乘車公里數的標準差爲20公里, 問

1. 隨機抽取300位乘客爲一組樣本,其其平均乘車公里數介於52及55公里間的機率爲何?

2. 續題1, 平均乘車公里數超過60公里的機率爲何?

圖9.21　台鐵乘客平均乘車公里數的機率



Solution:

1. 令 $\overline{X}$ 表示乘客平均乘車公里數, 因樣本夠大(n=300>30) 所以根據中央極限定理, 平均乘車公里數爲常態分配, 其平均數與標準差爲

$$\mu_{\overline{X}} = \mu = 55.1$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{300}} = 1.15$$

$$P(52 < \overline{X} < 55) = P(\frac{52-55.1}{1.15} < Z < \frac{55-55.1}{1.15}) = P(-2.70 < Z < -0.09) = 0.4606$$

2.

$$P(\overline{X} > 60) = P(Z > \frac{60-55.1}{1.15}) = P(Z > 4.26) = 0$$

9.7 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**

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

林惠玲　陳正倉著 雙葉書廊發行 2004

圖9.22 品質管制圖

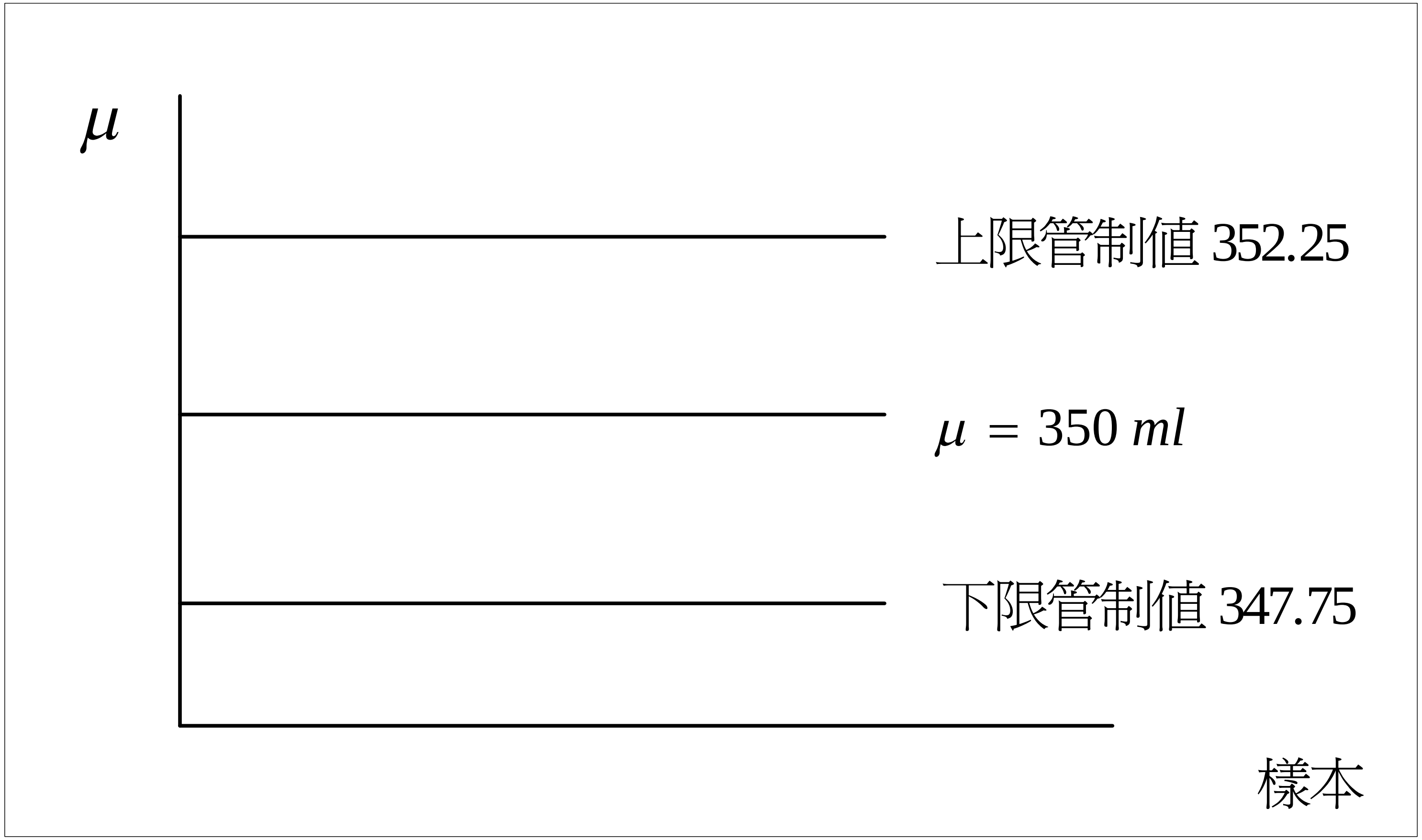
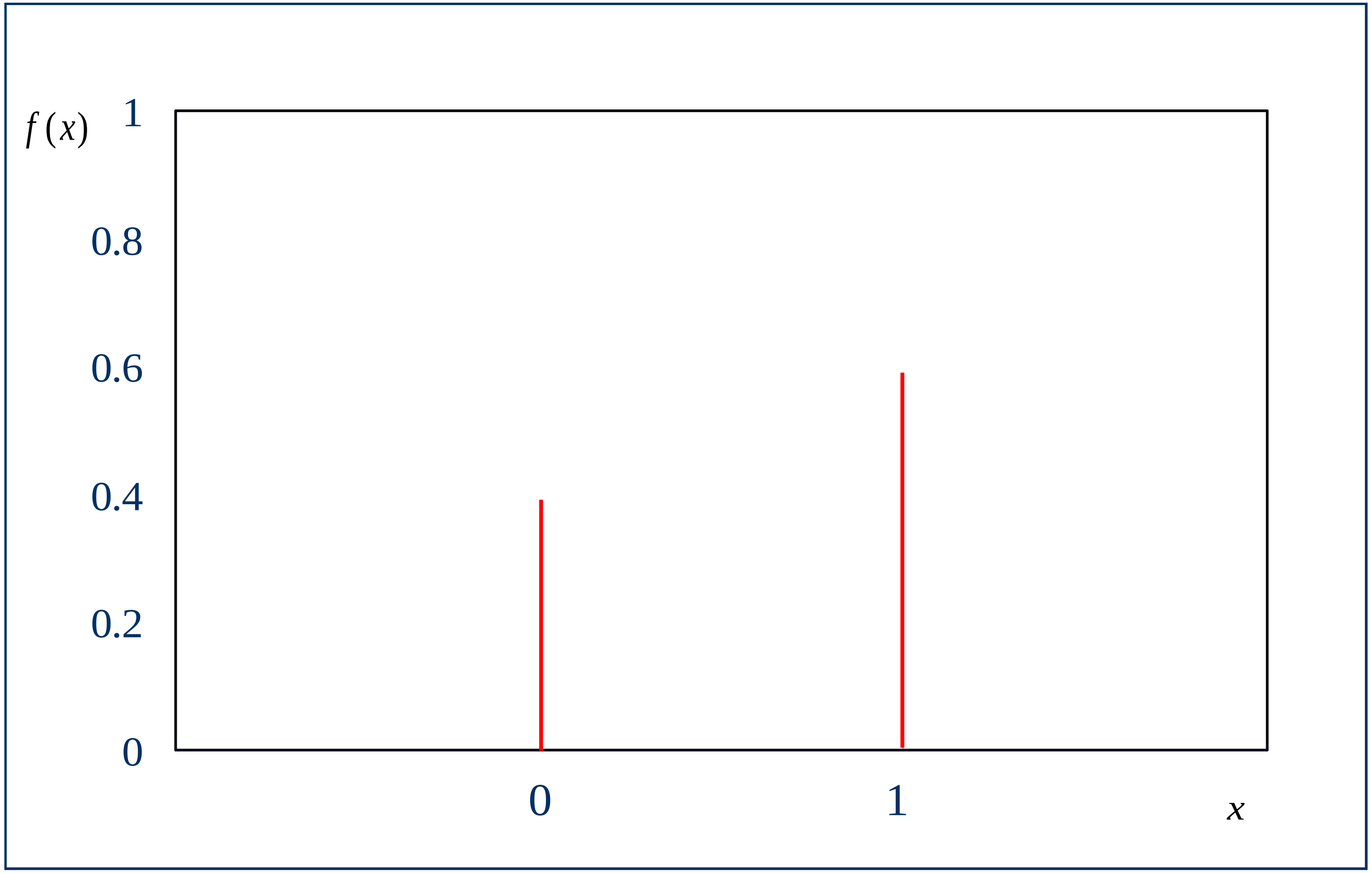


圖9.23 點二項分配



9.7.2 樣本比例的抽樣分配

樣本比例爲自點二項母體抽出的樣本平均數, 該樣本比例的機率分配表爲 $f(\hat{p})$,即爲樣本比例的抽樣分配

林惠玲 陳正倉著 雙葉書廊發行 2004

表9.13　母體比例

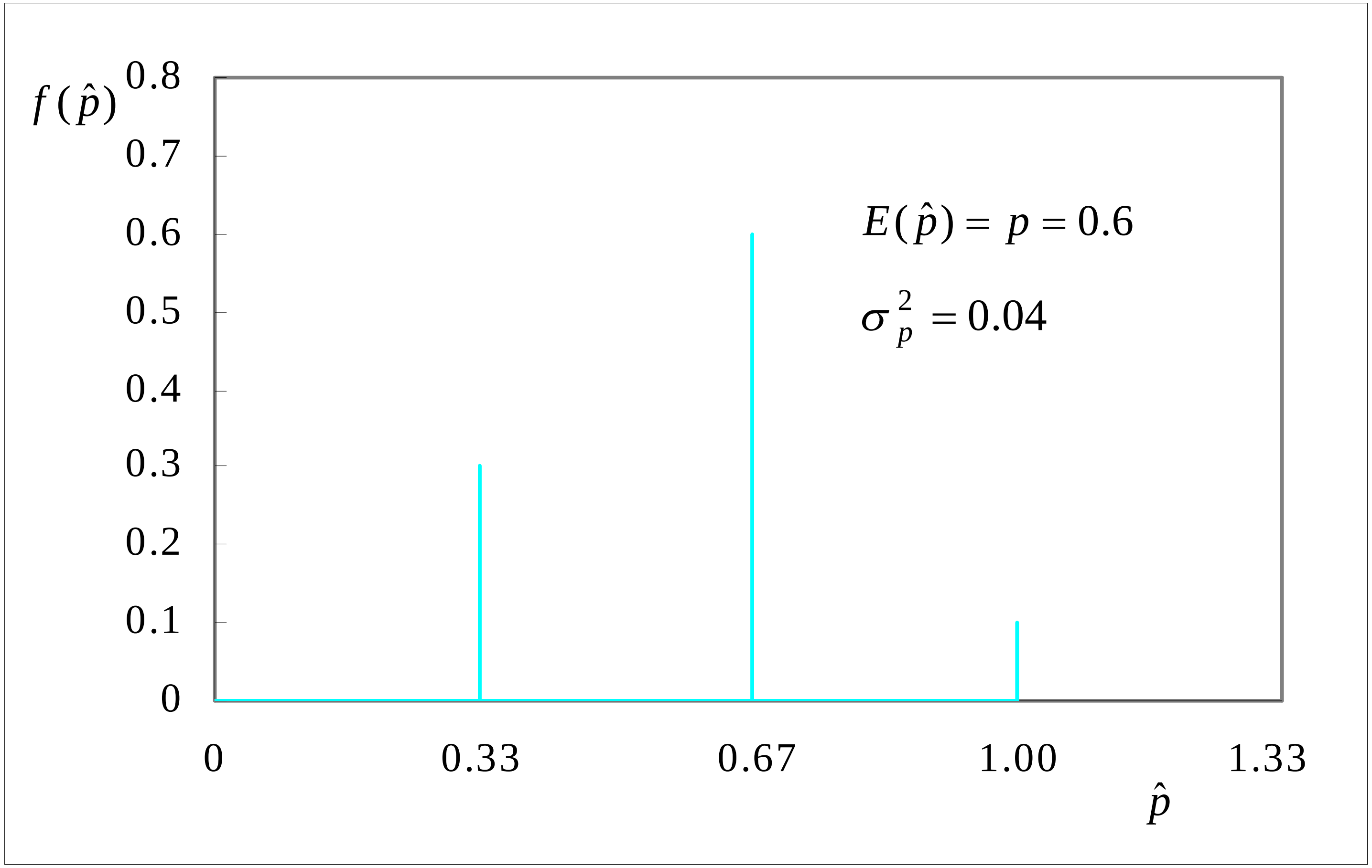
姓氏	林小姐	洪小姐	戴小姐	葉小姐	呂小姐
修讀會計學	讀過Y	沒讀過N	讀過Y	沒讀過N	讀過Y

林惠玲　陳正倉著 雙葉書廊發行 2004

表9.14　樣本比例的抽樣分配

$\hat{p}$	$f(\hat{p})$
0.33	3/10=0.3
0.67	6/10=0.6
1.00	1/10=0.1
	$\sum f(\hat{p})=1.00$

圖9.24　樣本比例的抽樣分配



林惠玲　陳正倉著 雙葉書廊發行 2004

9.7.3 樣本比例的平均數與變異數

❍ 樣本比例的變異數與標準差

①無限母體

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

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

②有限母體

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

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

林惠玲　陳正倉著 雙葉書廊發行 2004

9.7.4 樣本比例抽樣分配的形狀

大樣本 (np>5及 nq>5)樣本比例的抽樣分配

依據中央極限定理,當大樣本時, $\hat{p}$ 的抽樣分配會趨近於常態分配

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

小樣本比例的抽樣分配

無限母體 $\hat{p}$ ~ 二項分配 (p, pq/n)

有限母體 (n/N >=0.05)

$\hat{p}$ ~ 超幾何分配 $p, \frac{pq}{n} \cdot \frac{N-n}{N-1}$

林惠玲 陳正倉著 雙葉書廊發行 2004

9.7.5 樣本比例抽樣分配的應用

See EX 9.18

圖9.25 樣本平均數的抽樣

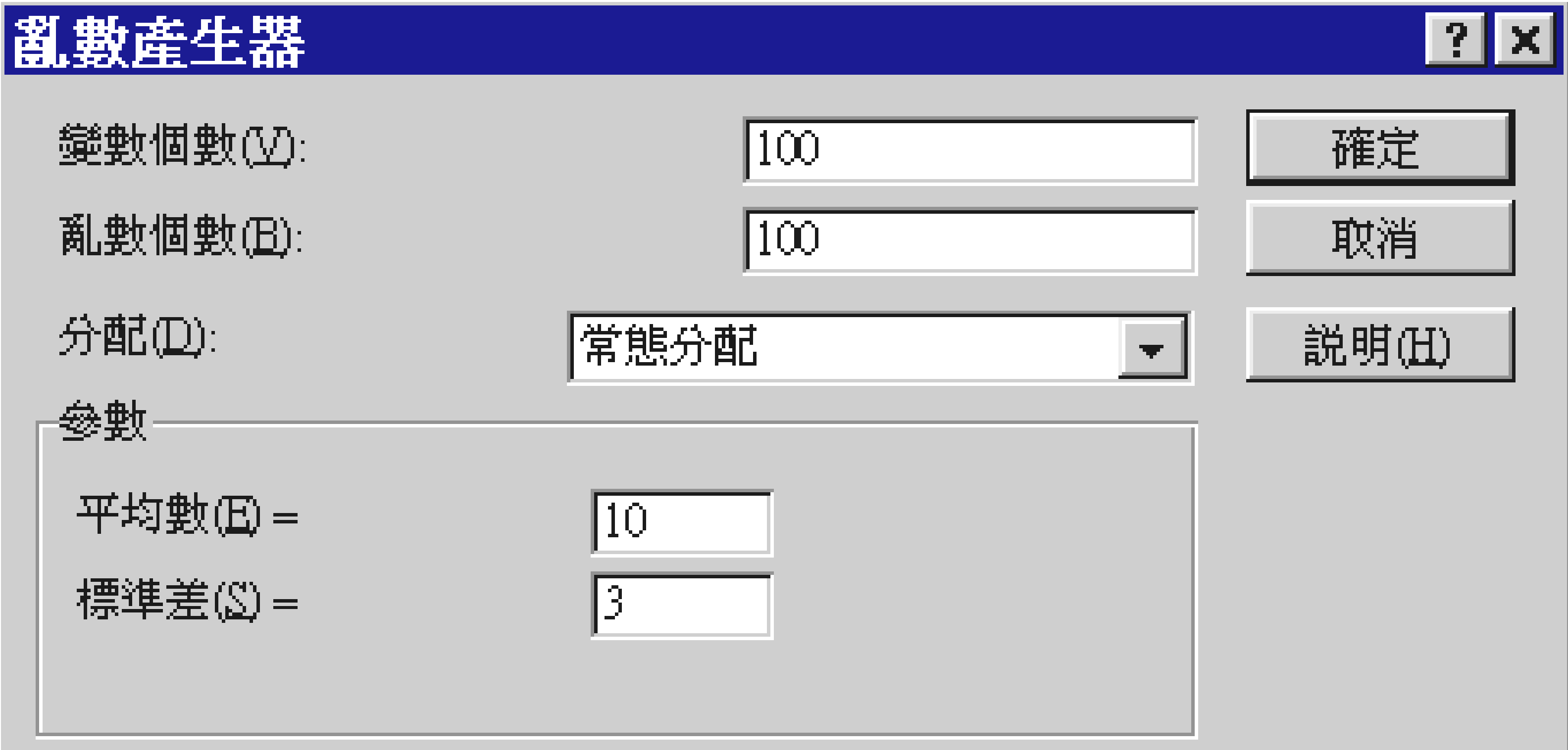


表9.15 100 組樣本平均數

	CS	CT	CU	CV	CW	CX
1	7.480235	7.536615	8.713022	8.639915	9.878545	9.4
2	10.17647	13.79036	10.78657	10.17141	10.53305	9.65
3	5.541916	5.81149	9.568558	8.113117	10.13049	9.9
4	8.580128	9.285656	6.705724	10.22454	10.17575	10.15
5	10.44954	7.124314	12.52598	9.477448	9.946502	10.4
6	10.58699	13.45946	13.04985	9.679591	10.5188	10.65
7	13.03795	8.313186	15.05604	7.102432	10.24558	11
8	6.486893	7.922102	11.85789	13.15603	9.596052	

林惠玲 陳正倉著 雙葉書廊發行 2004

表9.16 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	

圖9.26 直方圖

組界	頻率
9.4	4
9.65	12
9.9	24
10.15	31
10.4	15
10.65	12
11	2