七、抽樣與抽樣分布

Chapter 7 Sampling and sampling distribution

目錄

學習目標

知識(認知)

1.可以描述抽樣和抽樣分布的意涵。

2.可以說明各種抽樣分布之期望值和變異數的意涵。

3.分辨各種母體屬性與條件下，抽樣分布的差異性。

4.評價抽樣的使用價值。

技能

1.能夠計算各種母體屬性與條件下，樣本平均值抽樣分布的期望值和變異數。

2.能夠計算各種母體屬性與條件下，樣本比例抽樣分布的期望值和變異數。

3.綜合所學，能夠計算實務領域中各種情境下，樣本平均值或樣本比率抽樣分布的期望值和變異數。

態度(情意)

1.意識到在實務應用上，抽樣對資料蒐集和資料價值的重要性。

2.能夠判斷在各種情境下，抽樣分布的使用價值。

教學使用時間：5 小時

在實際地調查或研究應用領域上，皆沒有足夠的資源對母體中所有的基本單位進行調查或研究，因此，僅能透過抽樣的方法，對一部份基本單位所構成的組合進行調查或研究，獲得資訊，透過分析，以取得母體的資訊。

透過適當的抽樣方法可以獲得有代表性的樣本，分析此樣本的數值，可以得到有價值的樣本統計值(sample statistics)，進而提出正確的統計推論。凸顯抽樣程序在統計推論的重要性。為了瞭解由樣本統計值(statistics)推測母體參數(parameter)過程中，由抽樣所產生的誤差，因此需要對樣本統計值的抽樣分布有更深入的認識。

從母體資料求得平均值 μ、比率、變方 σ^2 或標準(偏)差 σ 在統計學上稱為**母數**、**介量**或**參數**(parameter)，由**樣本資料**求得之平均值$\bar{x}$、比率、樣本變異數 S^2 或標準(偏)差 S，在統計學上稱為**統計值**、**統計量**、**介值**(statistic)、**樣本統計量**(sample statistic)或**樣本統計值**。

樣本統計值屬於隨機變數，其數值會隨樣本不同而異。

在經營一家餐廳或旅館時，經常面臨碰到抽樣調查產品品質、服務品質、消費者滿意度、消費者意見等，不確定性的問題與挑戰，欲衡量其發生的可能性大小，就必須具體地善用抽樣分布的概念，清楚的呈現不確定性問題的可能性高低，以便於管理者採取適當的方式因應。

推論統計學(Inferential statistics)：又稱統計推論學(Statistical inference)或歸納統計學(Inductive statistics)

主要根據樣本統計值(statistic)對全部母體參數(parameter)進行統計推論。

一般性學術性研究或市場調查，受限時間、人力、物力、財力等因素，無法進行普查(全部研究對象納入調查研究)，只能抽取樣本(一部份基本單位所構成的組合)資料進行研究和調查分析。

推論統計又可分為有母數統計學(parametric statistics)和無母數統計學(non-parametric statistics)兩大領域。

有母數方法(parametric method)

母體是屬於**常態分布**或**樣本中基本單位數量 *n* 大**時，利用樣本統計值(statistic)進行推估與假設檢定。

利用常態分布(分配)、*t*-分布、卡方分布和 *F*-分布進行母體參數(parameter)之估計與假設檢定。

無母數方法(nonparametric method)或**自由分配法**(distribution-free method)

母體分布未知、**非常態分布母體**或**小樣本**條件進行統計推估，利用樣本資料之大小、順序或等級的特性進行統計推論。

統計推論(Statistical Inference)：包含假設檢定與統計估計兩種：

假設檢定(hypothesis testing / tests of hypothesis)

利用樣本統計值判定接受或拒絕研究者對母體參數的研究假設之方法。例：學生中不同性別的消費金額有顯著性差異？學生中不同性別的花心程度有顯著性差異？

統計估計(statistic estimation)

利用樣本統計值(statistic)數值推估母體參數(parameter)(高低、大小、多少或範圍)的方法。例：此餐廳有多少消費者願意再度蒞臨？此餐廳消費者消費金額的分布？

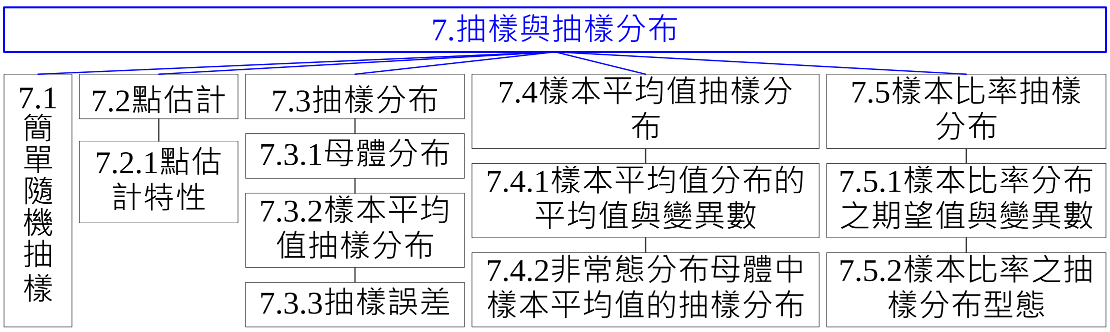
統計估計分為點估計(point estimation)和區間估計(interval estimation)兩種：

點估計(Point estimation)

利用抽樣所獲得的樣本統計值(statistic)進行母體特定研究調查參數(parameter)的數值之估算。

區間估計(Interval estimation)

利用抽樣所獲得的樣本統計值(statistic)進行母體特定研究調查參數的可信區間之估算。



章節結構圖

7.1 簡單隨機抽樣

在隨機抽樣過程中，抽取母體中每一個基本單位被抽中機率皆獨立且相等。藉由**簡單隨機抽樣(Simple random sampling)**或**單純隨機抽樣法**所獲得的樣本稱為簡單隨機樣本(simple random sample)或隨機樣本。受到母體基本單位數量 N 的情況，簡單隨機抽樣可以分為有限母體基本單位數量和無限母體基本單位數量兩種情況。

有限(finite)母體基本單位數量簡單隨機抽樣

從 N 個基本單位所構成的母體中，抽出 n 個基本單位(構成的組合)為樣本，若是簡單隨機抽樣，其**每一個基本單位**被**抽中機率**必須**相同**。

彩球法

先將母體 N 個基本單位編號，若母體數量 N 少於 1000 時，而該抽樣程序又要經常執行時，可以採用彩球法，製作 N 個彩球或彩券，依序編號，將彩球或彩券隨機置於同一個籃子內，在抽取 n 個彩球或彩券，抽取 n 次過程中抽取彩球或彩券後不放回原來籃子，抽中的彩球或彩券編號與基本單位編號相同者，該基本單位即為樣本。

亂數表法

先將母體 N 個基本單位編號，若母體達到千位，建議使用四位數亂數表(random number table)，母體達百位，建議使用四位數亂數表的前三位或後三位當成三位數亂數表。從亂數表中第一列或第一欄的數值開始，當亂數表內數值等於或小於 N 時，該亂數表數值標號的基本單位即為樣本；若亂數表內數值大於 N 時，該數值跳過不計；若亂數表內後面數值與前面數值重複時，該後面數值跳過不計。依序進行，達到原先設定的樣本數即可停止。

軟體抽樣法 SPSS 和 Excel

先將母體 N 個基本單位編號後，輸入相關數值統計軟體，依據軟體指令設定抽樣數量，即可獲得簡單隨機抽樣的樣本。

無限(infinite)母體基本單位數量近似簡單隨機抽樣

母體基本單位數量 N 可能是無窮大、很難精確計算或母體數量是動態(變動)。

無限母體基本單位數量的簡單隨機抽樣必須確定樣本中每一個基本單位皆來自相同的母體和樣本中每一個基本單位皆獨立抽出且機率相等。

例如：小琦餐廳欲調查中午用餐時間(1100~1400)的顧客對點餐人員服務的滿意度，先設定欲達到樣本的數量與比率，設欲取大約$\frac{1}{20}$比例的顧客當成樣本，取 1 個彩球或彩券標示為樣本，取 19 個彩球或彩券標示為非樣本，共 20 個彩球或彩券放入一個籃子，當顧客點餐完畢時，給顧客抽取一個彩球或彩券。若抽到標示為樣本的彩球或彩券，該位顧客列為樣本，若抽到標示為非樣本的彩球或彩券，該位顧客為非樣本。每位顧客抽取彩球或彩券後，不管結果為何，皆將彩球或彩券放回籃子中，供下一位顧客重新抽取。

Excel 函數

利用 Excel 軟體插入(I)→函數(F)…→選擇函數類別：**數學與三角函數**中選取 RAND()函數，即會產生大於等於 0 且小於 1 的隨機亂數。

練習 7.1 有限基本單位數量的母體中，有 N = 200 個基本單位，試利用下列四位數亂數表數值，選出 6 個簡單隨機樣本。

6054	2739	8509	9139	9657	1986	3856	7930	9625	7171
6336	2174	9303	5295	6326	3107	1263	2925	7647	5041
9928	5663	1559	6954	4005	6482	6899	1278	4237	3912
2927	3388	3102	7228	0213	4755	4264	9867	2253	5658
6612	3040	2643	4076	1552	3023	1764	8304	7388	0173

答案：從前三位數值取得：198, 126, 155, 127, 021, 176；從後三位數值取得：054, 139, 171, 174, 107, 041

7.2 點估計

從母體基本單位中隨機抽出樣本，由樣本獲得的統計值(statistic)以作為(提供)母體參數(parameter)之估計值(estimate)。欲估計母體平均值 μ，以樣本平均值$\bar{x}$當作母體平均值 μ 的估計值(estimate)，此樣本平均值$\bar{x}$即為點估計(Point estimation)。同理，欲估計母體標準(偏)差 σ，以樣本標準(偏)差 S 當作母體標準(偏)差 σ 的估計值(estimate)，此樣本標準(偏)差 S 即為點估計。

練習 7.2　雯雯速食餐廳櫃臺服務員接受每位消費者點餐的服務時間，利用簡單隨機抽樣獲得下列 12 個隨機樣本(單位：秒)。(a)試計算櫃臺服務員接受每位消費者點餐服務時間的平均值之估計值；(b)試計算櫃臺服務員接受每位消費者點餐服務時間的標準(偏)差之估計值？(答案有效位數四捨五入取到小數點後第 2 位)

12.5	25.3	35.2	62.2	35.7	45.2
54.3	36.8	46.5	25.8	72.5	82.5

練習 7.3　雯雯速食餐廳櫃臺服務員接受每位消費者點餐的服務時間，利用簡單隨機抽樣獲得下列 6 個隨機樣本(單位：秒)。(a)試計算櫃臺服務員接受每位消費者點餐服務時間的平均值之估計值；(b)試計算櫃臺服務員接受每位消費者點餐服務時間的標準(偏)差之估計值？(答案有效位數四捨五入取到小數點後第 2 位)

54.3	36.8	46.5	25.8	72.5	82.5

題解：

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
54.3	1.2333	1.5211
36.8	-16.2667	264.6044
46.5	-6.5667	43.1211
25.8	-27.2667	743.4711
72.5	19.4333	377.6544
82.5	29.4333	866.3211
318.4		2296.6933

樣本平均值$\bar{x}=\sum_{i=1}^{n}\frac{x_i}{n}=\frac{54.3+36.8+46.5+25.8+72.5+82.5}{6}=\frac{318.4}{6}=53.0667$ 秒

樣本標準(偏)差 $S=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{2296.6933}{6-1}}=\sqrt{459.3387}=21.4319$ 秒

答案：(a)母體平均值點估計值 53.07 秒；(b)母體標準(偏)差點估計值 21.43 秒

7.2.1 點估計特性

7.2.1.1 一致性

樣本基本單位數量 n 愈多↑，其樣本平均值$\bar{x}$愈接近母體平均值 μ，故當樣本平均值$\bar{x}$愈接近母體平均值 μ 時，稱樣本平均值$\bar{x}$與母體平均值 μ 之間具有一致性(Consistency)；當樣本基本單位數量 $n\to\infty$(無窮

多)時，會導致樣本平均值$\bar{x}$ = 母體平均值 μ，此時兩者數值會呈現一致。故樣本基本單位數量 n 要夠大，才具有母體的代表性。

當樣本基本單位數量 n 愈多時，計算獲得的統計值即愈接近參數的真正數值。

7.2.1.2 無偏性

所有可能的樣本統計值(點估計值)的期望值皆會等於母體參數，即顯示其**無偏性**或**不偏性**(Unbiasedness)。故，所有可能樣本平均值$\bar{x}$的期望值等於母體平均值 μ 時，稱此樣本平均值$\bar{x}$為母體平均值 μ 的無偏估值(unbiased estimate, unbiased estimator)：

$$E(\bar{x}) = \mu$$

若對很多不同的基本單位所構成的樣本組合，在每一組樣本組合中分別計算特定的樣本統計值$\bar{x}$，此些樣本統計值$\bar{x}$的算術平均值會接近於母體平均值(參數) μ 的真正數值。

7.2.1.3 有效性

費雪(Fisher)將變方 σ^2 的倒數($\frac{1}{\sigma^2}$)稱為**精確度**(accuracy)，故變方(分散程度)愈大者其精確度愈小。精確度大之估計值稱為高效估值(efficient estimate)，高效估值較具有母體的代表性。

當有不同的樣本組合時，可以提供母體參數的數個點估計值，當該樣本組合中的變異數較小者，其估算的樣本統計值(statistic)比較接近母體參數(parameter)，即代表點估計值的**有效性**或**高效性**(Efficiency)。

7.2.1.4 充分性

利用樣本數值資料可計算出多種統計值(statistic)，最能代表母體參數的樣本統計值，即為具有充分性(Sufficiency)的統計值。

若樣本統計值(statistic)可以充分利用每一個樣本的觀測值去估算母體參數(parameter)時，稱該樣本統計值具有充分性。

7.3 抽樣分布

隨機抽樣樣本數量 n 的隨機樣本($x_1, x_2, x_3,\ldots, x_n$)，構成樣本資料的統計值、統計量或介值(statistic)，因此樣本統計值(statistic)為簡單隨機樣本的函數，樣本統計值亦屬於隨機變數，其機率分布稱為抽樣分布(sampling distribution)。

7.3.1 母體分布

母體分布(Population distribution)是母體中所有基本單位資料的機率分布。母體分布可為常態分布(Normal distribution)、二項分布(Binominal probability distribution)或一致性機率分布(uniform probability distribution)。

範例 7.1　小雯海產店過去五個月每月營業額(單位：新台幣元)(比例尺度)與盈餘(Earnings, E)或虧損(Loss, L)(名目尺度)的分布如下表所示：

月份	x_1	x_2	x_3	x_4	x_5
營業額	245000	255000	255000	265000	255000
盈餘或虧損	L	L	E	E	E

分別對營業額和盈餘或虧損，進行母體分布的分析

母體中營業額次數分布和機率分布

X	f	$f(x)$
245000	1	0.2
255000	3	0.6
265000	1	0.2
合計	5	1.0

營業額平均值和標準(偏)差

月份	x_i	$x_i - \mu$	$(x_i - \mu)^2$
1	245000	-10000	100000000
2	255000	0	0
3	255000	0	0
4	265000	10000	100000000
5	255000	0	0
合計	1275000		200000000

母體平均值 $\mu = \frac{245000+255000+255000+265000+255000}{5} = \frac{1275000}{5} = 255000$ 元

母體變異數 $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{200000000}{5} = 40000000$ 元 ^{2}

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{200000000}{5}} = \sqrt{40000000} = 6324.5$ 元

母體中盈餘(Earnings, E)或虧損(Loss, L)次數分布和機率分布

X	f	$f(x)$
盈餘(Earnings, E)	3	0.6
虧損(Loss, L)	2	0.4
合計	5	1.0

盈餘(Earnings, E)的期望值與標準(偏)差

在 n 次獨立事件(隨機實驗)中，盈餘機率為 p，虧損機率為 $q = 1 - p$，發生盈餘次數 X 的期望值 $E(X) = \mu = n \times p = 5 \times 0.6 = 3.0$

母體變異數 $Var(X) = \sigma^2 = n \times p \times q = n \times p \times (1 - p) = 5 \times 0.6 \times 0.4 = 1.2$

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{n \times p \times q} = \sqrt{n \times p \times (1-p)} = \sqrt{1.2} = 1.0954$

7.3.2 樣本平均值抽樣分布

母體中 N 個基本單位之隨機變數 X，機率分布 $f(x)$，從母體中隨機抽取 n 個基本單位為樣本，分別為 $x_1, x_2, x_3, \ldots, x_n$，樣本平均值$\bar{x} = \sum_{i=1}^{n}\frac{x_i}{n}$。所有可能的樣本平均值$\bar{x}$之機率分布 $f(\bar{x})$，即為樣本平均值$\bar{x}$的抽樣分布(sampling distribution)。

獲得樣本平均值$\bar{x}$之抽樣分布 $f(\bar{x})$的目的，運用於推估母體平均值 μ 實際數值。

從母體中全部 N 個基本單位隨機抽出 n 個基本單位為樣本，會產生C_n^N種樣本的組合。

$$C_n^N = \binom{N}{n} = \frac{N!}{n! \times (N-n)!} = {}_N C_n$$

範例 7.2　小雯海產店過去五個月每月營業額(單位：新台幣元)與盈餘(Earnings, E)或虧損(Loss, L)的分布如下表：

月份	x_1	x_2	x_3	x_4	x_5
營業額	245000	255000	255000	265000	255000
盈餘或虧損	L	L	E	E	E

分別對營業額和盈餘或虧損，進行樣本平均值抽樣分布的分析

題解：

若從中抽取 3 個月的營業額當樣本，會有可能的樣本組合為

$$C_3^5 = \binom{5}{3} = \frac{5!}{3!\times(5-3)!} = \frac{5\times4\times3\times2\times1}{3\times2\times1\times2\times1} = 10 \text{ 種}$$

可能的樣本組合分別為

樣本組合	營業額組合			$\bar{x}$
x_1, x_2, x_3	245000	255000	255000	251667
x_1, x_2, x_4	245000	255000	265000	255000
x_1, x_2, x_5	245000	255000	255000	251667
x_1, x_3, x_4	245000	255000	265000	255000
x_1, x_3, x_5	245000	255000	255000	251667
x_1, x_4, x_5	245000	265000	255000	255000
x_2, x_3, x_4	255000	255000	265000	258333
x_2, x_3, x_5	255000	255000	255000	255000
x_2, x_4, x_5	255000	265000	255000	258333
x_3, x_4, x_5	255000	265000	255000	258333

營業額樣本平均值$\bar{x}$的抽樣分布

$\bar{x}_i$	f	$f(\bar{x}_i)$
251667	3	0.3
255000	4	0.4
258333	3	0.3
合計	10	1.0

透過樣本平均值$\bar{x}$的抽樣分布可以推估，若某次抽樣獲得樣本營業額平均值$\bar{x}$為 251667 機率為 0.3，同理獲得樣本營業額平均值$\bar{x}$等於 255000 機率為 0.4。

樣本平均值$\bar{x}$抽樣分布的期望值(平均值) $E(\bar{x})$

$$E(\bar{x}) = \sum_{i=1}^{k} \bar{x}_i \times f(\bar{x}_i) = 251667 \times 0.3 + 255000 \times 0.4 + 258333 \times 0.3 = 255000$$

樣本組合	營業額組合			$\bar{x}_i$	$\bar{x}_i - E(\bar{x})$	$[\bar{x}_i - E(\bar{x})]^2$
x_1, x_2, x_3	245000	255000	255000	251666.67	-3333.33	11111111.11
x_1, x_2, x_4	245000	255000	265000	255000.00	0.00	0.00
x_1, x_2, x_5	245000	255000	255000	251666.67	-3333.33	11111111.11
x_1, x_3, x_4	245000	255000	265000	255000.00	0.00	0.00
x_1, x_3, x_5	245000	255000	255000	251666.67	-3333.33	11111111.11
x_1, x_4, x_5	245000	265000	255000	255000.00	0.00	0.00
x_2, x_3, x_4	255000	255000	265000	258333.33	3333.33	11111111.11
x_2, x_3, x_5	255000	255000	255000	255000.00	0.00	0.00
x_2, x_4, x_5	255000	265000	255000	258333.33	3333.33	11111111.11
x_3, x_4, x_5	255000	265000	255000	258333.33	3333.33	11111111.11
合計				2550000.00	0.00	66666666.67

樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}}$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k}[\bar{x}_i - E(\bar{x})]^2}{N}} = \sqrt{\frac{66666667}{10}} = \sqrt{6666666.7} = 2581.99$$

樣本中盈餘(Earnings, E)或虧損(Loss, L)次數分布和機率分布

若從中抽取 3 個月的營業額當樣本，會有可能的樣本組合為

$$C_3^5 = \binom{5}{3} = \frac{5!}{3!\times(5-3)!} = \frac{5\times4\times3\times2\times1}{3\times2\times1\times2\times1} = 10 \text{ 種}$$

可能的樣本組合分別為

樣本組合	營業額組合			$\bar{x}$	盈餘或虧損組合	盈餘比例$\bar{p}$
x_1, x_2, x_3	245000	255000	255000	251667	L, L, E	0.3333
x_1, x_2, x_4	245000	255000	265000	255000	L, L, E	0.3333
x_1, x_2, x_5	245000	255000	255000	251667	L, L, E	0.3333
x_1, x_3, x_4	245000	255000	265000	255000	L, E, E	0.6667
x_1, x_3, x_5	245000	255000	255000	251667	L, E, E	0.6667
x_1, x_4, x_5	245000	265000	255000	255000	L, E, E	0.6667
x_2, x_3, x_4	255000	255000	265000	258333	L, E, E	0.6667
x_2, x_3, x_5	255000	255000	255000	255000	L, E, E	0.6667
x_2, x_4, x_5	255000	265000	255000	258333	L, E, E	0.6667
x_3, x_4, x_5	255000	265000	255000	258333	E, E, E	1.0000

盈餘或虧損樣本比例$\bar{p}$的抽樣分布

$\bar{p}$	f	$f(x)$
0.3333	3	0.3
0.6667	6	0.6
1.0000	1	0.1
合計	10	1.0

盈餘的樣本比例$\bar{p}$與標準(偏)差

樣本組合	營業額組合			$\bar{x}$	盈餘或虧損組合	盈餘比例$\bar{p}_i$	$\bar{p}_i - E(\bar{p})$	$[\bar{p}_i - E(\bar{p})]^2$
x_1, x_2, x_3	245000	255000	255000	251667	L, L, E	0.3333	-0.2667	0.0711
x_1, x_2, x_4	245000	255000	265000	255000	L, L, E	0.3333	-0.2667	0.0711
x_1, x_2, x_5	245000	255000	255000	251667	L, L, E	0.3333	-0.2667	0.0711
x_1, x_3, x_4	245000	255000	265000	255000	L, E, E	0.6667	0.0667	0.0044
x_1, x_3, x_5	245000	255000	255000	251667	L, E, E	0.6667	0.0667	0.0044
x_1, x_4, x_5	245000	265000	255000	255000	L, E, E	0.6667	0.0667	0.0044
x_2, x_3, x_4	255000	255000	265000	258333	L, E, E	0.6667	0.0667	0.0044
x_2, x_3, x_5	255000	255000	255000	255000	L, E, E	0.6667	0.0667	0.0044
x_2, x_4, x_5	255000	265000	255000	258333	L, E, E	0.6667	0.0667	0.0044
x_3, x_4, x_5	255000	265000	255000	258333	E, E, E	1.0000	0.4000	0.1600
合計						6.0000		0.4001

盈餘的樣本比例$\bar{p}$抽樣分布的期望值 $E(\bar{p}) = \frac{\sum_{i=1}^{n}\bar{p}_i}{n} = \frac{6.0000}{10} = 0.6$

盈餘的樣本比例$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}} = \sqrt{\frac{\sum_{i=1}^{k=10}[\bar{p}_i - E(\bar{p})]^2}{N}} = \sqrt{\frac{0.4001}{10}} = \sqrt{0.0400} = 0.20$$

練習 7.4　一個摸彩箱中有三個等質球，分別標示 1、2 和 3 數值，每一個球被抽中機率獨立和相等。以抽出後不放回的方式，隨機抽出 2 球為樣本，請估算(A)此隨機試驗的樣本空間；(B)樣本平均值$\bar{x}$的抽樣分布；(C)樣本平均值$\bar{x}$的期望值與變異數(答案有效位數四捨五入取到小數點後第 4 位)。

題解：(A)樣本空間 $S = \{(1,2), (1,3), (2,1), (2,3), (3,1), (3,2)\}$

(B)各種樣本組合之平均值與其機率表

樣本組合	$\bar{x}$	機率$f(\bar{x})$
1,2	1.5	1/6 = 0.1667
1,3	2.0	1/6 = 0.1667
2,1	1.5	1/6 = 0.1667
2,3	2.5	1/6 = 0.1667
3,1	2.0	1/6 = 0.1667
3,2	2.5	1/6 = 0.1667

樣本平均值$\bar{x}$的抽樣分布

$\bar{x}_i$	機率$f(\bar{x}_i)$
1.5	1/3 = 0.3333
2.0	1/3 = 0.3333
2.5	1/3 = 0.3333

(C)樣本平均值$\bar{x}$抽樣分布的期望值(平均值) $E(\bar{x})$

$$E(\bar{x}) = \sum_{i=1}^{k} \bar{x}_i \times f(\bar{x}_i) = 1.5\times0.3333 + 2.0\times0.3333 + 2.5\times0.3333 = 2.0$$

樣本平均值$\bar{x}$抽樣分布的母體變異數$\sigma_{\bar{x}}^2$

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{k}[\bar{x}_i - E(\bar{x})]^2}{N} = \frac{(1.5-2.0)^2+(2.0-2.0)^2+(2.5-2.0)^2}{3} = \frac{0.25+0+0.25}{3} = \frac{0.5}{3} = \frac{1}{6} = 0.1667$$

答案：(A)樣本空間 S = {(1,2), (1,3), (2,1), (2,3), (3,1), (3,2)}；(C)$E(\bar{x})$ = 2.0；$\sigma_{\bar{x}}^2 = \frac{1}{6}$ = 0.1667

練習 7.5　從雯雯速食餐廳四位櫃臺服務員(2 男 2 女)中隨機抽出 2 位服務員，透過此 2 位服務員昨天的銷售金額，推估四位櫃臺服務員的昨天平均銷售金額。若昨天四位櫃臺服務員(2 男 2 女依序)銷售金額分別為新台幣 6000、6500、7500 和 7000 元。請估算(A)採用簡單隨機抽樣，所有可能之樣本為何，請列出；(B)採用依據性別比例配置之分層隨機抽樣法，所有可能之樣本為何，請列出；(C)請計算(A)中樣本平均值$\bar{x}$的抽樣分布；(D)在(A)中樣本平均值$\bar{x}$的期望值與變異數(答案有效位數四捨五入取到小數點後第 4 位)。

7.3.3 抽樣誤差

母體中 N 個基本單位之隨機變數 X，機率分布 $f(x)$，從母體中隨機抽取 n 個基本單位為樣本，分別為 $x_1, x_2, x_3, \ldots, x_n$，樣本平均值$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$。若從母體 N 個基本單位隨機抽出 n 個基本單位為樣本，會產生C_n^N種樣本的組合。每一種樣本組合的樣本平均值$\bar{x}$可能會有差異。

樣本平均值$\bar{x}$(統計值)與母體平均值 μ(母體參數)的差異量即可視為抽樣誤差(sampling error)。

$$抽樣誤差 = |\bar{x} - \mu|$$

範例 7.3　小雯海產店過去 5 個月每月營業額(單位：新台幣元)與盈餘(Earnings, E)或虧損(Loss, L)的分布如下表所示。(a)母體基本單位數量 $N = 5$，隨機抽取 $n = 2$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(b)母體基本單位數量 $N = 5$，隨機抽取 $n = 3$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(c)母體基本單位數量 $N = 5$，隨機抽取 $n = 4$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(答案有效位數四捨五入取到小數點後第 4 位)

月份	x_1	x_2	x_3	x_4	x_5
營業額	245000	255000	255000	265000	255000
盈餘或虧損	L	L	E	E	E

題解：

(a)母體基本單位數量 $N = 5$，隨機抽取 $n = 2$ 個基本單位為樣本，會產生可能的樣本組合為

$$C_2^5 = \binom{5}{2} = \frac{5!}{2!\times(5-2)!} = \frac{5\times4\times3\times2\times1}{2\times1\times3\times2\times1} = 10 \text{ 種}$$

可能的樣本組合

樣本組合	營業額組合		$\bar{x}$	$\lvert\bar{x} - \mu\rvert$
x_1, x_2	245000	255000	250000	5000
x_1, x_3	245000	255000	250000	5000
x_1, x_4	245000	265000	255000	0
x_1, x_5	245000	255000	250000	5000
x_2, x_3	255000	255000	255000	0
x_2, x_4	255000	265000	260000	5000
x_2, x_5	255000	255000	255000	0
x_3, x_4	255000	265000	260000	5000
x_3, x_5	255000	255000	255000	0
x_4, x_5	265000	255000	260000	5000

母體平均值 $\mu = \frac{245000+255000+255000+265000+255000}{5} = \frac{1275000}{5} = 255000$

沒有產生抽樣誤差機率 $p = \frac{4}{10} = 0.4000$

(b)母體基本單位數量 $N = 5$，隨機抽取 $n = 3$ 個基本單位為樣本，會產生可能的樣本組合為

$$C_3^5 = \binom{5}{3} = \frac{5!}{3!\times(5-3)!} = \frac{5\times4\times3\times2\times1}{3\times2\times1\times2\times1} = 10 \text{ 種}$$

可能的樣本組合

樣本組合	營業額組合			$\bar{x}$	$\lvert\bar{x} - \mu\rvert$
x_1, x_2, x_3	245000	255000	255000	251667	3333
x_1, x_2, x_4	245000	255000	265000	255000	0
x_1, x_2, x_5	245000	255000	255000	251667	3333
x_1, x_3, x_4	245000	255000	265000	255000	0
x_1, x_3, x_5	245000	255000	255000	251667	3333
x_1, x_4, x_5	245000	265000	255000	255000	0
x_2, x_3, x_4	255000	255000	265000	258333	3333
x_2, x_3, x_5	255000	255000	255000	255000	0
x_2, x_4, x_5	255000	265000	255000	258333	3333
x_3, x_4, x_5	255000	265000	255000	258333	3333

在每一種抽樣組合中，大都會產生抽樣誤差，透過樣本平均值的抽樣分布，可以推估在所有的抽樣組合中，可能產生抽樣誤差機率。例如：產生 3333 抽樣誤差機率為 0.6，沒有產生抽樣誤差機率為 0.4。

沒有產生抽樣誤差機率 $p = \frac{4}{10} = 0.4000$

(c)母體基本單位數量 $N = 5$，隨機抽取 $n = 4$ 個基本單位為樣本，會產生可能的樣本組合為

$$C_4^5 = \binom{5}{4} = \frac{5!}{4!\times(5-4)!} = \frac{5\times4\times3\times2\times1}{4\times3\times2\times1\times1} = 5 \text{ 種}$$

可能的樣本組合

樣本組合	營業額組合				$\bar{x}$	$\lvert\bar{x}-\mu\rvert$
x_1, x_2, x_3, x_4	245000	255000	255000	265000	255000	0
x_1, x_2, x_3, x_5	245000	255000	255000	255000	252500	2500
x_1, x_2, x_4, x_5	245000	255000	265000	255000	255000	0
x_1, x_3, x_4, x_5	245000	255000	265000	255000	255000	0
x_2, x_3, x_4, x_5	255000	255000	265000	255000	257500	2500

沒有產生抽樣誤差機率 $p = \frac{3}{5} = 0.6000$

答案：(a) 0.4000；(b) 0.4000；(c) 0.6000

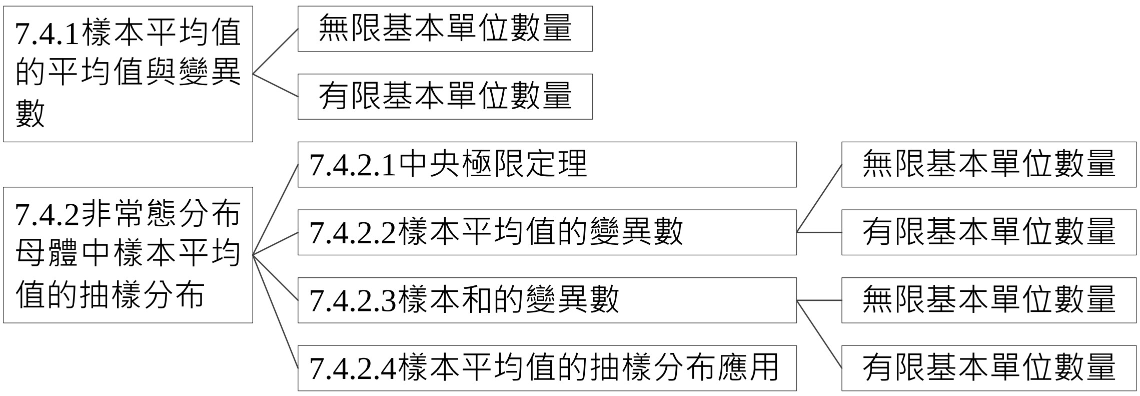
練習 7.6　小鈺海產店過去 6 個月每月營業額(單位：新台幣元)與盈餘(Earnings, E)或虧損(Loss, L)的分布如下表所示。(a)母體基本單位數量 $N = 6$，隨機抽取 $n = 2$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(b)母體基本單位數量 $N = 6$，隨機抽取 $n = 3$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(c)母體基本單位數量 $N = 6$，隨機抽取 $n = 4$ 個基本單位為樣本，請計算抽出樣本平均值，沒有產生抽樣誤差機率？(答案有效位數四捨五入取到小數點後第 4 位)

月份	x_1	x_2	x_3	x_4	x_5	x_6
營業額	245000	255000	255000	265000	255000	250000
盈餘或虧損	L	L	E	E	E	E

題解：

由以上分析可知，母體基本單位數量 N 愈多，基本單位數值分散程度愈高(標準(偏)差愈高)，沒有產生抽樣誤差機率愈低，會產生抽樣誤差機率愈高。

7.4 樣本平均值抽樣分布



7.4.1 樣本平均值抽樣分布的平均值與變異數

母體中 N 個基本單位之隨機變數 X，機率分布 $f(x)$，從母體中隨機抽取 n 個基本單位為樣本，分別為 $x_1, x_2, x_3, \ldots, x_n$，樣本平均值 $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$。若從母體 N 個基本單位隨機抽出 n 個基本單位為樣本，會產生 C_n^N 種樣本的組合。

C_n^N種樣本組合之樣本平均值$\bar{x}$分布的母體平均值，利用$\mu_{\bar{x}}$或 $E(\bar{x})$符號表示，即$\mu_{\bar{x}} = E(\bar{x}) = \mu$。

$$\mu_{\bar{x}} = E(\bar{x}) = E(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n} \times E(\sum_{i=1}^{n} x_i) = \frac{1}{n} \times E(x_1 + x_2 + x_3 + \ldots + x_n) = \frac{1}{n} \times [E(x_1) + E(x_2) + E(x_3) + \ldots + E(x_n)] = \frac{1}{n} \times (\mu + \mu + \mu + \ldots + \mu) = \frac{1}{n} \times n \times \mu = \mu$$

故在樣本平均值$\bar{x}$抽樣分布中的母體平均值$\mu_{\bar{x}}$等於原始資料隨機變數 X 分布的母體平均值 μ。

C_n^N種樣本組合之樣本平均值$\bar{x}$分布的母體變異數，利用$\sigma_{\bar{x}}^2$、$Var(\bar{x})$或 $V(\bar{x})$表示；C_n^N種樣本組合之樣本平均值$\bar{x}$分布的母體標準(偏)差(standard deviation of the sampling distribution of the sample mean)，利用$\sigma_{\bar{x}}$符號表示。其中，σ^2 為原始資料隨機變數 X 的母體變異數；σ 為原始資料隨機變數 X 的母體標準(偏)差。

$$\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = V(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n^2} \times V(\sum_{i=1}^{n} x_i) = \frac{1}{n^2} \times V(x_1 + x_2 + x_3 + \ldots + x_n) = \frac{1}{n^2} \times [V(x_1) + V(x_2) + V(x_3) + \ldots + V(x_n)] = \frac{1}{n^2} \times (\sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

樣本平均值$\bar{x}$抽樣分布的母體變異數和標準(偏)差計算方式會受母體的型態分為無限或有限而異。

無限基本單位數量 N 之母體、樣本比例低($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，供下一個樣本抽取)：

$$\text{樣本平均值}\bar{x}\text{抽樣分布的母體變異數}\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = \frac{\sigma^2}{n}$$

$$\text{樣本平均值}\bar{x}\text{抽樣分布的母體標準(偏)差}\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

有限基本單位數量 N 之母體：

$$\text{樣本平均值}\bar{x}\text{抽樣分布的母體變異數}\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$

$$\text{樣本平均值}\bar{x}\text{抽樣分布的母體標準(偏)差}\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為有限母體校正因子(finite population correction factor)。在有限的基本單位數量下的母體，當抽出的樣本(基本單位)不放回原來的母體中，以供下一個樣本的抽樣時，先抽到的樣本之機率與後抽到樣本的機率會有差異，特別在母體總數 N 不多的情況愈明顯。因此，若母體基本單位總數量 N 不多的情況下，抽到的樣本又不放回原先的母體中，以供下一個樣本抽樣時，就必須進行有限母體校正。

一般在抽出樣本數量 n 高於母體總數量 N 的 5 %以上時($\frac{n}{N} > 0.05$)，計算樣本平均值$\bar{x}$分布的母體變異數$\sigma_{\bar{x}}^2$和母體標準(偏)差$\sigma_{\bar{x}}$才需要校正。若抽出樣本數量低於母體總數量的 5 %以下時($\frac{n}{N} \le 0.05$)，計算樣本平均值分布的母體變異數和母體標準(偏)差不需要校正，計算方式與無限基本單位數量之母體相同。

範例 7.4　大王海產店過去五個月每月營業額(單位：新台幣元)與盈餘(Earnings, E)或虧損(Loss, L)的分布如下表：

月份	x_1	x_2	x_3	x_4	x_5
營業額	245000	255000	255000	265000	255000
盈餘或虧損	L	L	E	E	E

試估算隨機抽取 3 個月資料樣本平均值$\bar{x}$抽樣分布的變異數$\sigma_{\bar{x}}^2$和標準(偏)差$\sigma_{\bar{x}}$。

題解：

依據抽樣分布的狀況，一一列出樣本組合運算方式。若從中抽取 3 個月的營業額當樣本，會產生可能的樣本組合數量為

$$C_3^5 = \binom{5}{3} = \frac{5!}{3! \times (5-3)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = 10 \text{ 種}$$

所有可能的樣本組合

樣本組合	營業額組合			$\bar{x}$
x_1, x_2, x_3	245000	255000	255000	251667
x_1, x_2, x_4	245000	255000	265000	255000
x_1, x_2, x_5	245000	255000	255000	251667
x_1, x_3, x_4	245000	255000	265000	255000
x_1, x_3, x_5	245000	255000	255000	251667
x_1, x_4, x_5	245000	265000	255000	255000
x_2, x_3, x_4	255000	255000	265000	258333
x_2, x_3, x_5	255000	255000	255000	255000
x_2, x_4, x_5	255000	265000	255000	258333
x_3, x_4, x_5	255000	265000	255000	258333

營業額樣本平均值$\bar{x}$的抽樣分布

$\bar{x}_i$	f	$f(\bar{x}_i)$
251667	3	0.3
255000	4	0.4
258333	3	0.3
合計	10	1.0

透過樣本平均值$\bar{x}$的抽樣分布可以推估，若某次抽樣獲得樣本營業額平均值$\bar{x}$為 251667 機率為 0.3，同理獲得樣本營業額平均值$\bar{x}$等於 255000 機率為 0.4。

樣本平均值$\bar{x}$抽樣分布的期望值(平均值) $E(\bar{x})$

$$E(\bar{x}) = \sum_{i=1}^{k} \bar{x}_i \times f(\bar{x}_i) = 251667 \times 0.3 + 255000 \times 0.4 + 258333 \times 0.3 = 255000$$

樣本組合	營業額組合			$\bar{x}_i$	$\bar{x}_i - E(\bar{x})$	$[\bar{x}_i - E(\bar{x})]^2$
x_1, x_2, x_3	245000	255000	255000	251667	-3333	11111111
x_1, x_2, x_4	245000	255000	265000	255000	0	0
x_1, x_2, x_5	245000	255000	255000	251667	-3333	11111111
x_1, x_3, x_4	245000	255000	265000	255000	0	0
x_1, x_3, x_5	245000	255000	255000	251667	-3333	11111111
x_1, x_4, x_5	245000	265000	255000	255000	0	0
x_2, x_3, x_4	255000	255000	265000	258333	3333	11111111
x_2, x_3, x_5	255000	255000	255000	255000	0	0
x_2, x_4, x_5	255000	265000	255000	258333	3333	11111111
x_3, x_4, x_5	255000	265000	255000	258333	3333	11111111
合計				2550000		66666667

樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k}[\bar{x}_i - E(\bar{x})]^2}{N}} = \sqrt{\frac{66666667}{10}} = \sqrt{6666666.7} = 2581.989$

利用有限母體修正計算法，運算樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}}$。原始資料隨機變數 X 分布的母體變異數 $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N} = \frac{200000000}{5} = 40000000$。

樣本平均值$\bar{x}$抽樣分布的母體變異數$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \times \frac{N-n}{N-1} = \frac{40000000}{3} \times \frac{5-3}{5-1} = 6666667$

樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \sqrt{6666667} = 2581.989$ 兩種算法數值都是相同

答案：樣本平均值抽樣分布之母體變異數$\sigma_{\bar{x}}^2 = 6666667$；樣本平均值抽樣分布之母體標準(偏)差$\sigma_{\bar{x}} = 2581.989$

7.4.2 非常態分布母體中樣本平均值的抽樣分布

7.4.2.1 中央極限定理

在任何型態的分布(包括非常態分布和常態分布)中，設母體平均值為 μ，標準(偏)差為 σ(μ 及 σ 皆為有限數值)，從母體中隨機抽取 n 個觀測值為樣本，所有可能組合之樣本平均值$\bar{x}$分布的母體平均值($\mu_{\bar{x}}$)等於原始資料隨機變數 X 母體平均值(μ)，即$\mu_{\bar{x}} = \mu$ 。

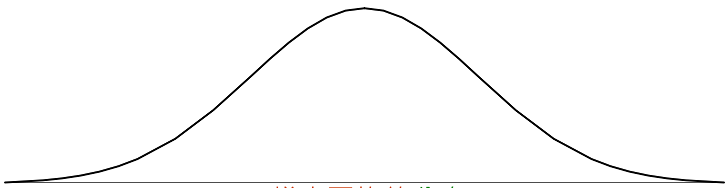
> 母體平均值為 μ，標準(偏)差為 σ(μ 及 σ 皆為有限數值)的母體中，當樣本數量愈大($n \geq 30$)時，樣本平均值$\bar{x}$的抽樣分布愈接近(趨近、近似)常態分布，且向隨機變數 X 母體平均值 μ 之位置集中。

在中央極限定理(Central limit theorem, CLT)成立的前提下，樣本平均值$\bar{x}$的抽樣分布會趨近於常態分布$\bar{x}\sim N(\mu, \frac{\sigma^2}{n})$，故可使用於推估特定樣本平均值$\bar{x}$出現機率。

樣本平均值$\bar{x}$標準化值的分布 $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$，當原始資料隨機變數 X 分布的母體標準(偏)差 σ 已知，計算樣本平均值$\bar{x}$所對應的 Z 值。

樣本平均值$\bar{x}$標準化值的分布 $Z = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} \sim N(0, 1)$，當原始資料隨機變數 X 分布的母體標準(偏)差 σ 未知，計算樣本平均值$\bar{x}$所對應的 Z 值。

加入圖形協助陳述



樣本平均值分布

範例 7.5　以 2、4 和 6 三個數(基本單位)組成母體，若三個出現的次數一樣多(機率相同，抽出後又放回，以供下一個樣本抽取)，其原始分布屬於矩形分布(rectangular distribution)。利用抽樣數量的增加說明中央極限定理。

題解：**第一部分**：若每次從母體中隨機抽取 1 個基本單位為樣本($n = 1$)，此樣本之分布仍為矩形分布。

第二部分：若每次從母體中隨機抽取 2 個基本單位為樣本($n = 2$)，可能有的樣本組合為：

樣本		樣本平均值$\bar{x}$
2	2	2
2	4	3
2	6	4
4	2	3
4	4	4
4	6	5
6	2	4
6	4	5
6	6	6

樣本數量 $n = 2$ 時，樣本平均值$\bar{x}$機率分布

樣本平均值$\bar{x}$ ($n = 2$)	次數 f	$f(x)$
2	1	0.1111
3	2	0.2222
4	3	0.3333
5	2	0.2222
6	1	0.1111

第三部分：若每次從母體抽取 3 個體為樣本($n = 3$)，可能有的樣本組合為：

樣本	樣本平均值$\bar{x}$	樣本	樣本平均值$\bar{x}$	樣本	樣本平均值$\bar{x}$
2 2 2	2.0000	4 2 2	2.6667	6 2 2	3.3333
2 2 4	2.6667	4 2 4	3.3333	6 2 4	4.0000
2 2 6	3.3333	4 2 6	4.0000	6 2 6	4.6667
2 4 2	2.6667	4 4 2	3.3333	6 4 2	4.0000
2 4 4	3.3333	4 4 4	4.0000	6 4 4	4.6667
2 4 6	4.0000	4 4 6	4.6667	6 4 6	5.3333
2 6 2	3.3333	4 6 2	4.0000	6 6 2	4.6667
2 6 4	4.0000	4 6 4	4.6667	6 6 4	5.3333
2 6 6	4.6667	4 6 6	5.3333	6 6 6	6.0000

樣本數量 $n = 3$ 時，樣本平均值$\bar{x}$機率分布

樣本平均值$\bar{x}$ ($n = 3$)	次數 f	$f(x)$
2.0000	1	0.0370
2.6667	3	0.1111
3.3333	6	0.2222
4.0000	7	0.2593
4.6667	6	0.2222
5.3333	3	0.1111
6.0000	1	0.0370

第四部分：樣本數量 $n = 4$ 時，樣本平均值$\bar{x}$機率分布，顯示樣本平均值的分數會以原始資料隨機變數 X 分布的母體平均值為中心，左右對稱(表格中顯示上下對稱)的特性。

樣本平均值$\bar{x}$ ($n = 4$)	次數 f	$f(x)$
2.0	1	0.0123
2.5	4	0.0494
3.0	10	0.1235
3.5	16	0.1975
4.0	19	0.2346
4.5	16	0.1975
5.0	10	0.1235
5.5	4	0.0494
6.0	1	0.0123

原始隨機變數 X 母體觀測值分布雖非常態分布，只要樣品數量 n 夠大，其樣品平均值$\bar{x}$所組成的新母體分布會接近常態分布。

7.4.2.2 樣本平均值抽樣分布的變異數與標準(偏)差

在任何型態的分布(包括非常態分布和常態分布)中，設母體平均值為 μ，標準(偏)差為 σ(μ 及 σ 皆為有限數值)，從母體中隨機抽取 n 個觀測值為樣本，所有可能組合之樣本平均值$\bar{x}$抽樣分布的平均值($\mu_{\bar{x}}$)等於隨機變數 X 的母體平均值(μ)，即$\mu_{\bar{x}} = \mu$。

無限基本單位數量之母體：

樣本平均值$\bar{x}$抽樣分布的母體變異數$\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = \frac{\sigma^2}{n}$

樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

有限基本單位數量之母體：

樣本平均值$\bar{x}$抽樣分布的母體變異數$\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$

樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為**有限母體校正因子**(finite population correction factor)。

範例 7.6　假設母體平均值 $\mu = 80$，標準(偏)差 $\sigma = 10$ 和基本單位數量 1000 個母體中，利用簡單隨機抽樣抽取 250 個隨機樣本，抽樣後樣本不放回。試估算(A)樣本平均值抽樣分布的期望值；(B)樣本平均值抽樣分布的標準(偏)差？(答案有效位數四捨五入取到小數點後第 2 位)

題解：母體平均值 $\mu = 80$，母體標準(偏)差 $\sigma = 10$，母體基本單位數量 $N = 1000$，樣本數量 $n = 250$。

樣本平均值抽樣分布的期望值$\mu_{\bar{x}} = \mu = 80.00$

樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{10}{\sqrt{250}} \times \sqrt{\frac{1000-250}{1000-1}} = 0.5480$

答案：(A)樣本平均值抽樣分布的期望值 80.00；(B)樣本平均值抽樣分布的母體標準(偏)差 0.55

練習 7.7　假設母體平均值 $\mu = 20$，標準(偏)差 $\sigma = 12$ 和基本單位數量 2000 個母體中，利用簡單隨機抽樣抽取 300 個隨機樣本，抽樣後樣本不放回。試估算(A)樣本平均值抽樣分布的期望值；(B)樣本平均值抽樣分布的標準(偏)差？(答案有效位數四捨五入取到小數點後第 2 位)

題解：母體平均值 $\mu = 20$，母體標準(偏)差 $\sigma = 12$，母體基本單位數量 $N = 2000$，樣本數量 $n = 300$。

樣本平均值抽樣分布的期望值$\mu_{\bar{x}} = \mu = 20.00$

樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{12}{\sqrt{300}} \times \sqrt{\frac{2000-300}{2000-1}} = 0.6389$

答案：(A)樣本平均值抽樣分布的期望值 20.00；(B)樣本平均值抽樣分布的母體標準(偏)差 0.64

表 7-1　樣本平均值$\bar{x}$之抽樣分布

母體屬性	條件	樣本平均值抽樣分布
常態分布 $E(X) = \mu$ $Var(X) = \frac{\sigma^2}{n}$	無限基本單位數量之母體、$\frac{n}{N} \le 0.05$ 或有限基本單位數量之母體(抽出的樣本放回) $n \ge 30$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$
	有限基本單位數量之母體或$\frac{n}{N} > 0.05$ (抽出的樣本不放回) $n < 30$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$
非常態分布 $E(X)$ $Var(X)$	無限基本單位數量之母體、$\frac{n}{N} \le 0.05$ 或有限基本單位數量之母體(抽出的樣本放回) $n \ge 30$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$
	有限基本單位數量之母體或$\frac{n}{N} > 0.05$ (抽出的樣本不放回) $n < 30$	$\bar{x}$~母體分布模式

範例 7.7　台灣外食人口中，每人每週外食消費金額為常態分布(Normal distribution)，平均值 NT$ 800 元，標準(偏)差 NT$ 200 元，在隨機抽樣中獲得樣本中每人每週消費金額之平均值$\bar{x}$。計算樣本數量 n 為 10、100 和 1000 時，每人每週消費金額之平均值$\bar{x}$抽樣分布的平均值$\mu_{\bar{x}}$和標準(偏)差$\sigma_{\bar{x}}$。

題解：隨機變數 X 代表每人每週外食消費金額，母體平均值 μ = NT$ 800 元，母體標準(偏)差 σ = NT$ 200 元。

樣本數量 $n = 10$：樣本平均值抽樣分布的平均值$\mu_{\bar{x}} = \mu$ = NT$ 800 元，樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{200}{\sqrt{10}} = \frac{200}{3.1623}$ = NT$ 63.2 元

樣本數量 $n = 100$：樣本平均值抽樣分布的平均值$\mu_{\bar{x}} = \mu$ = NT\$ 800 元，樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{200}{\sqrt{100}} = \frac{200}{10}$ = NT\$ 20.0 元

樣本數量 $n = 1000$：樣本平均值抽樣分布的平均值$\mu_{\bar{x}} = \mu$ = NT\$ 800 元，樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{200}{\sqrt{1000}} = \frac{200}{31.6228}$ = NT\$ 6.3 元

答案：平均值皆為 NT\$ 800 元；$n = 10$、100 和 1000 依序分別的母體標準(偏)差為 NT\$ 63.2、20.0 和 6.3 元

故當樣本數 n 愈多時，消費金額平均值$\bar{x}$抽樣分布的標準(偏)差$\sigma_{\bar{x}}$愈小，代表當樣本數 n 增加時，樣本平均值$\bar{x}$抽樣分布的分散程度會減少，愈往原始資料隨機變數 X 分布的母體平均值 μ 數值集中。

練習 7.8　由常態分布的母體 $N(16, 4^2)$中，隨機抽出假設 X_1、X_2、X_3、...、X_{15} 的一組樣本數量 15 個樣本。請估計(A)$E(\bar{x})$；(B)$V(\bar{x})$；(C)$P(14.32 < \bar{x} < 16.34)$；(D)$P(13.32 < \bar{x} < 16.00)$。(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體 $N(16, 4^2)$平均值=16；變異數 $\sigma^2 = 4^2$；標準差 $\sigma = \sqrt{\sigma^2} = \sqrt{4^2} = 4$

(A)樣本平均值抽樣分布的期望值 $E(\bar{x}) = \mu_{\bar{x}} = \mu = 16.0000$

(B)樣本平均值抽樣分布的母體變異數 $V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{4^2}{15} = 1.0667$

(C)$P(14.32 < \bar{x} < 16.34) = P(\frac{14.32-\mu}{\sigma_{\bar{x}}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{16.34-\mu}{\sigma_{\bar{x}}}) = P(\frac{14.32-16.00}{\sqrt{1.0667}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{16.34-16.00}{\sqrt{1.0667}}) = P(-1.6267 \le Z \le 0.3292)$

$= P(Z \le 0.3292) - P(Z \le -1.6267) = 0.6290 - 0.0519$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.5771

(D)$P(13.32 < \bar{x} < 16.00) = P(\frac{13.32-\mu}{\sigma_{\bar{x}}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{16.00-\mu}{\sigma_{\bar{x}}}) = P(\frac{13.32-16.00}{\sqrt{1.0667}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{16.00-16.00}{\sqrt{1.0667}}) = P(-2.5949 \le Z \le 0.0000)$

$= P(Z \le 0.0000) - P(Z \le -2.5949) = 0.5000 - 0.0047$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.4953

答案：(A) 16.0000；(B) 1.0667；(C) 0.5771；(D) 0.4953

練習 7.9　已知義式咖啡飲料體積的標準(偏)差 σ = 2.4 ml，其平均值 μ 未知，現隨機抽取 90 杯義式咖啡飲料，其體積的平均值為$\bar{x}$，請計算$|\bar{x} - \mu| < 0.4$ ml 機率。(答案有效位數四捨五入取到小數點後第 4 位)

題解：因為樣本數量 $n = 90 > 30$ 屬於大量樣本，所以依據中央極限定理樣本平均值$\bar{x} \sim N[E(\bar{x}), V(\bar{x})]$，樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{2.4}{\sqrt{90}} = \frac{2.4}{9.4868}$ = 0.252982 ml

$P(|\bar{x} - \mu| < 0.4) = P(-0.4 < \bar{x} - \mu < 0.4) = P(\frac{-0.4}{\sigma_{\bar{x}}} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{0.4}{\sigma_{\bar{x}}}) = P(\frac{-0.4}{0.252982} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{0.4}{0.252982}) = P(-1.581139 < Z < 1.581139) = P(Z \le 1.581139) - P(Z \le -1.581139) = 0.943077 - 0.056923$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.8862

答案：$P(|\bar{x} - \mu| < 0.4) = 0.8862$

範例 7.8　已知小琦咖啡店每天平均販售 360 杯卡布其諾咖啡飲料，其標準(偏)差為 15 杯，若以 35 天為一樣本，利用$\bar{x}$代表每天平均販售的卡布其諾杯數。(A)請陳述$\bar{x}$的分布狀況；(B) $\bar{x}$介於 358 到 365 杯機率；(C)一天販售卡布其諾超過 370 杯機率。(答案有效位數四捨五入取到小數點後第 4 位)

題解：隨機變數 X 代表一天販售卡布其諾咖啡飲料杯數。

(A)因為樣本數量 $n = 35 > 30$ 屬於大量樣本，所以依據中央極限定理樣本平均值$\bar{x} \sim N[E(\bar{x}), V(\bar{x})]$。樣本平均值$\bar{x}$抽樣分布的期望值$\mu_{\bar{x}} = \mu$ = 360 杯，樣本平均值$\bar{x}$抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{35}} = \frac{15}{5.91608}$ = 2.5355 杯。偏態係數 $\beta_1 = 0$，峰態係數 $\beta_2 = 0$。

(B)$P(358 \le \bar{x} \le 365) = P(\frac{358-360}{2.535463} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{365-360}{2.535463}) = P(-0.7888 \le Z \le 1.9720) = P(Z \le 1.9720) - P(Z \le -0.7888) =$ 0.9757 – 0.2151(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.7606

(C)$P(X \ge 370) = P(\frac{x-\mu}{\sigma} \ge \frac{370-360}{15}) = P(Z \ge 0.6667) = 1 - P(Z \le 0.6667) = 1 - 0.7475$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.2525

答案：(A)樣本平均值$\bar{x} \sim N[E(\bar{x}), V(\bar{x})]$，$\bar{x}$抽樣分布的期望值$\mu_{\bar{x}} = \mu = 360$ 杯，樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = 2.5355$ 杯，偏態係數 $\beta_1 = 0$，峰態係數 $\beta_2 = 0$；(B) $P(358 \le \bar{x} \le 365) = 0.7606$；(C) $P(X \ge 370) = 0.2525$

練習 7.10 小雯連鎖咖啡館男性與女性櫃臺服務員每小時銷售金額皆屬於常態分布，男性服務員銷售金額 $X \sim N(6490,300)$，女性服務員銷售金額 $Y \sim N(6500,200)$，隨機分別抽出 10 位男性與女性服務員其分別銷售金額的平均值為$\bar{x}$和$\bar{y}$。請計算(A)$P(\bar{y} > \bar{x})$；(B)$P(|\bar{y} - \bar{x}| > 10)$。(答案有效位數四捨五入取到小數點後第 4 位)

題解：隨機變數 X 代表男性銷售金額 $X \sim N(6490,$變異數 $= 300)$，樣本數量 $n = 10$，樣本平均值$\bar{x} \sim N(6490,$變異數 $= \frac{300}{10})$。隨機變數 Y 代表女性銷售金額 $Y \sim N(6500,$變異數 $= 200)$，樣本數量 $n = 10$，樣本平均值 $\bar{y} \sim N(6500,$變異數 $= \frac{200}{10})$。兩個獨立的常態分布變數相減後之分布$\bar{y} - \bar{x} \sim N(6500 - 6490,$變異數 $= \frac{300}{10} + \frac{200}{10})$ (平均值相減，變異數相加)。

(A) $P(\bar{y} > \bar{x}) = P(\bar{y} - \bar{x} > 0) = P(\frac{(\bar{y}-\bar{x})-(\mu_y-\mu_x)}{\sigma_{\bar{y}-\bar{x}}} > \frac{0-10}{\sqrt{\frac{300}{10}+\frac{200}{10}}}) = P(Z > \frac{-10}{7.0711}) = P(Z > -1.4142) = 1 - P(Z \le -1.4142) =$ 1 – 0.0787(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9214

(B) $P(|\bar{y} - \bar{x}| > 10) = P(\bar{y} - \bar{x} > 10) + P(\bar{y} - \bar{x} < -10) = P(\frac{(\bar{y}-\bar{x})-(\mu_y-\mu_x)}{\sigma_{\bar{y}-\bar{x}}} > \frac{10-10}{\sqrt{\frac{300}{10}+\frac{200}{10}}}) + P(\frac{(\bar{y}-\bar{x})-(\mu_y-\mu_x)}{\sigma_{\bar{y}-\bar{x}}} < \frac{-10-10}{\sqrt{\frac{300}{10}+\frac{200}{10}}}) =$ $P(Z > \frac{0}{7.0711}) + P(Z < \frac{-20}{7.0711}) = P(Z > 0) + P(Z < -2.8284) = 0.5 + 0.0023$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.5023

答案：(A) $P(\bar{y} > \bar{x}) = 0.9214$；(B) $P(|\bar{y} - \bar{x}| > 10) = 0.5023$

範例 7.9 甲與乙兩個母體皆屬於常態分布，標準(偏)差皆等於 3，$\mu_{甲} - \mu_{乙} = 2$。請估算(A)分別從甲與乙兩個母體各隨機抽出 1 個樣本，甲樣本的數值比乙樣本數值高機率；(B)分別從甲與乙兩個母體各隨機抽出 7 個樣本，甲樣本平均值比乙樣本平均值高機率；(C)每次接從甲與乙兩母體各隨機抽 1 個樣本，共抽 5 次，此 5 次甲樣本的數值皆比乙樣本數值高機率。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

(A) $P(甲 > 乙) = P(甲 - 乙 > 0) = P(\frac{(甲-乙)-\left(\mu_{甲}-\mu_{乙}\right)}{\sigma_{甲-乙}} > \frac{0-2}{\sqrt{3^2+3^2}}) = P(Z > \frac{-2}{4.2426}) = P(Z > -0.4714) = 1 - P(Z \le -0.4714) = 1 - 0.3187$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.6813

(B) $P(\overline{甲} > \overline{乙}) = P(\overline{甲} - \overline{乙} > 0) = P(\frac{\left(\overline{甲}-\overline{乙}\right)-\left(\mu_{甲}-\mu_{乙}\right)}{\sigma_{\overline{甲}-\overline{乙}}} > \frac{0-2}{\sqrt{\frac{3^2}{7}+\frac{3^2}{7}}}) = P(Z > \frac{-2}{1.6036}) = P(Z > -1.2472) = 1 - P(Z \le -1.2472) = 1 - 0.1062$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.8938

(C) $f(5) = \binom{n}{x} \times p^x \times q^{n-x} = \binom{5}{5} \times 0.6813^5 \times 0.3187^0 = 0.1468$

答案：(A) $P(甲 > 乙) = 0.6813$；(B) $P(\overline{甲} > \overline{乙}) = 0.8938$；(C) 0.1468

練習 7.11　100 年度四技二專聯招英文考科考生分數屬於常態分布，平均分數 80 分，標準(偏)差 20 分。請估算(A)若隨機選取 70 位考生，請計算其平均分數高於 82 分機率；(B)若該年度有 12 萬名考生，請計算分數高於 60 分的人數。(機率答案有效位數四捨五入取到小數點後第 4 位；人數答案取到個位數)

題解：隨機變數 X 代表考生分數 $X\sim N(80,20^2)$，樣本數量 $n = 70$，樣本平均值$\bar{x}\sim N(80,\frac{20^2}{70})$。

(A)$P(\bar{x} \geq 82) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \geq \frac{82-80}{\sqrt{\frac{20^2}{70}}}) = P(Z \geq \frac{2}{2.3905}) = P(Z \geq 0.8367) = 1 – P(Z \leq 0.8367) = 1 – 0.7986$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.2014

(B)$120000 \times P(X \geq 60) = 120000 \times P(\frac{x-\mu}{\sigma} \geq \frac{60-80}{20}) = 120000 \times P(Z \geq -1) = 120000 \times [1 – P(Z \leq -1)] = 120000 \times (1 – 0.1587) = 100961.4$ 人

答案：(A)$P(\bar{x} \geq 82) = 0.2014$；(B)高於 60 分有 100961 人

練習 7.12　100 年度四技二專聯招英文考科考生分數屬於常態分布，平均分數 80 分，標準(偏)差 20 分，隨機抽取樣本數量分別為 10 和 15 位考生為甲和乙兩組樣本。請估算(A)$P(|\bar{x}_{甲} - \bar{x}_{乙}| > 0.3)$；(B)$P(\bar{x}_{乙} - \bar{x}_{甲} > 0.5)$；(C)$P(\bar{x}_{甲} - \bar{x}_{乙} = 0.0)$；(D)$P(\bar{x}_{乙} - \bar{x}_{甲} = 0.3)$；(E)$P(\bar{x}_{甲} - \bar{x}_{乙} > 0.5)$。(答案有效位數四捨五入取到小數點後第 4 位)

題解：隨機變數 X 代表考生分數 $X\sim N(80,20^2)$，$n_{甲} = 10$，$\bar{x}_{甲}\sim N(80,\frac{20^2}{10})$，$n_{乙} = 15$，$\bar{x}_{乙}\sim N(80,\frac{20^2}{15})$。

(A) $\bar{x}_{甲} - \bar{x}_{乙}\sim N(80\text{-}80,\frac{20^2}{10}+\frac{20^2}{15})$

$P(|\bar{x}_{甲} - \bar{x}_{乙}| > 0.3) = P(\bar{x}_{甲} - \bar{x}_{乙} > 0.3) + P(\bar{x}_{甲} - \bar{x}_{乙} < –0.3) = P(\frac{(\bar{x}_{甲}-\bar{x}_{乙})-(\mu_{甲}-\mu_{乙})}{\sigma_{\bar{x}_{甲}-\bar{x}_{乙}}} > \frac{0.3-0}{\sqrt{\frac{20^2}{10}+\frac{20^2}{15}}}) + P(\frac{(\bar{x}_{甲}-\bar{x}_{乙})-(\mu_{甲}-\mu_{乙})}{\sigma_{\bar{x}_{甲}-\bar{x}_{乙}}} < \frac{-0.3-0}{\sqrt{\frac{20^2}{10}+\frac{20^2}{15}}}) = P(Z > \frac{0.3}{8.1650}) + P(Z < \frac{-0.3}{8.1650}) = P(Z > 0.0367) + P(Z < -0.0367) = 1 - P(Z \leq 0.0367) + P(Z < -0.0367) = 1 – 0.5147 + 0.4853$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9707

(B) $\bar{x}_{乙} - \bar{x}_{甲}\sim N(80\text{-}80,\frac{20^2}{10}+\frac{20^2}{15})$

$P(\bar{x}_{乙} - \bar{x}_{甲} > 0.5) = P(\frac{(\bar{x}_{乙}-\bar{x}_{甲})-(\mu_{乙}-\mu_{甲})}{\sigma_{\bar{x}_{乙}-\bar{x}_{甲}}} > \frac{0.5-0}{\sqrt{\frac{20^2}{10}+\frac{20^2}{15}}}) = P(Z > \frac{0.5}{8.1650}) = P(Z > 0.0612) = 1 – P(Z \leq 0.0612) = 1 – 0.5244$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.4756

(C) $\bar{x}_{甲} - \bar{x}_{乙}\sim N(80\text{-}80,\frac{20^2}{10}+\frac{20^2}{15})$

$P(\bar{x}_{甲} - \bar{x}_{乙} = 0.0) = 0.0000$，在連續機率分布中，特定一個點機率為 0.0000

(D)$P(\bar{x}_{乙} - \bar{x}_{甲} = 0.3) = 0.0000$，在連續機率分布中，特定一個點機率為 0.0000

(E)$P(\bar{x}_{甲} - \bar{x}_{乙} > 0.5) = P(\frac{(\bar{x}_{甲}-\bar{x}_{乙})-(\mu_{甲}-\mu_{乙})}{\sigma_{\bar{x}_{甲}-\bar{x}_{乙}}} > \frac{0.5-0}{\sqrt{\frac{20^2}{10}+\frac{20^2}{15}}}) = P(Z > \frac{0.5}{8.1650}) = P(Z > 0.0612) = 1 – P(Z \leq 0.0612) = 1 – 0.5244$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.4756

答案：(A) $P(|\bar{x}_{甲} - \bar{x}_{乙}| > 0.3) = 0.9707$；(B) $P(\bar{x}_{乙} - \bar{x}_{甲} > 0.5) = 0.4756$；(C) $P(\bar{x}_{甲} - \bar{x}_{乙} = 0.0) = 0.0000$；(D) $P(\bar{x}_{乙} - \bar{x}_{甲} = 0.3) = 0.0000$；(E)$P(\bar{x}_{甲} - \bar{x}_{乙} > 0.5) = 0.4756$

練習 7.13 有兩個常態分布的母體，分別為 $N(3, 12)$與 $N(5, 8)$，分別隨機抽出 4 個樣本，其平均值分別為 $\bar{x}_1$和$\bar{x}_2$。試估算(A)$P(\bar{x}_1 < \bar{x}_2)$機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：樣本 1 平均值$\bar{x}_1 \sim N(3, \frac{12}{4})$，樣本 2 平均值$\bar{x}_2 \sim N(5, \frac{8}{4})$，兩樣本平均值差的抽樣分布$\bar{x}_1 - \bar{x}_2 \sim N(3\text{-}5, \frac{12}{4} + \frac{8}{4})$

$P(\bar{x}_1 < \bar{x}_2) = P(\bar{x}_1 - \bar{x}_2 < 0) = P(\frac{(\bar{x}_1-\bar{x}_2)-(\mu_2-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} < \frac{0-(3-5)}{\sqrt{\frac{12}{4}+\frac{8}{4}}}) = P(Z < \frac{2}{2.2361}) = P(Z < 0.8944) = 0.8144$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

答案：(A) $P(\bar{x}_1 < \bar{x}_2) = 0.8144$

範例 7.10 一個常態分布的母體，其分布為 $N(\mu, 12)$，隨機抽出 10 個基本單位為樣本，其平均值為$\bar{x}$。試估算(A)$P(\mu - 2 \le \bar{x} \le \mu + 2)$；(B)$P(\bar{x} \le \mu + 1)$？(答案有效位數四捨五入取到小數點後第 4 位)

題解：樣本平均值$\bar{x}$抽樣分布$\sim N(\mu,$變異數 $= \frac{12}{10})$。

(A)$P(\mu - 2 \le \bar{x} \le \mu + 2) = P(\frac{(\mu-2)-\mu}{\sqrt{\frac{12}{10}}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{(\mu+2)-\mu}{\sqrt{\frac{12}{10}}}) = P(\frac{-2}{1.0954} \le Z \le \frac{2}{1.0954}) = P(-1.8257 \le Z \le 1.8257) = P(Z \le 1.8257) - P(Z \le -1.8257) = 0.9661 - 0.0339$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9321

(B)$P(\bar{x} \le \mu + 1) = P(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{(\mu+1)-\mu}{\sqrt{\frac{12}{10}}}) = P(Z \le \frac{1}{1.0954}) = P(Z \le 0.9129) = 0.8193$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

答案：(A)$P(\mu - 2 \le \bar{x} \le \mu + 2) = 0.9321$；(B)$P(\bar{x} \le \mu + 1) = 0.8193$

練習 7.14 深水大學學生每天消費金額屬於常態分布，現今從該校學生中隨機抽出 12 位學生。試估算(A)此 12 位學生與全部學生的平均消費金額之差不超過全部學生消費金額標準(偏)差的$\frac{1}{2}$之機率？(B)若該大學學生每天消費金額不屬於常態分布，請計算此 12 位學生與全部學生的平均消費金額之差不超過全部學生消費金額標準(偏)差的$\frac{1}{2}$之機率(答案有效位數四捨五入取到小數點後第 4 位)

題解：

(A) 樣本平均值$\bar{x}$抽樣分布$\sim N(\mu, \frac{\sigma^2}{12})$趨近於常態分布

$P(|\bar{x} - \mu| \le \frac{1}{2} \times \sigma) = P(|\frac{(\bar{x}-\mu)-(\mu-\mu)}{\sigma_{\bar{x}}}| \le \frac{\frac{1}{2}\times\sigma-0}{\sqrt{\frac{\sigma^2}{12}}}) = P(|Z| < \frac{\frac{\sigma}{2}}{\frac{\sigma}{3.4641}}) = P(|Z| < 1.7321) = P(Z < 1.7321) - P(Z < -1.7321) = 0.9584 - 0.0416$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9167

(B) $P(|\bar{x} - \mu| \le \frac{1}{2} \times \sigma) = P(|\bar{x} - \mu| \le k \times \sigma_{\bar{x}}) = P(|\bar{x} - \mu| \le k \times \frac{\sigma}{3.4641}) \ge 1 - \frac{1}{k^2} = 1 - \frac{1}{1.7321^2} = 0.6667$ (依據柴比氏定理)

$$\frac{1}{2} \times \sigma = k \times \frac{\sigma}{3.4641} \rightarrow \frac{1}{2} = k \times \frac{1}{3.4641} \rightarrow k = \frac{1}{2} \times 3.4641 = 1.7321$$

答案：(A) $P(|\bar{x} - \mu| \le \frac{1}{2} \times \sigma) = 0.9167$；(B) $P(|\bar{x} - \mu| \le \frac{1}{2} \times \sigma) = 0.6667$

7.4.2.3 樣本和抽樣分布的變異數與標準(偏)差【選擇教材】

在任何型態的分布(包括非常態分布和常態分布)中，母體中 N 個基本單位之隨機變數 X，設母體平均值為 μ，標準(偏)差為 σ(μ 及 σ 皆為有限數值)，從母體中隨機抽取 n 個基本單位為樣本，觀測值分別為 $x_1, x_2, x_3, \ldots, x_n$，其樣本和 $s = x_1 + x_2 + x_3 + \ldots + x_n = \sum_{i=1}^{n} x_i$。

樣本和抽樣分布的平均值$\bar{s} = E(s) = E(x_1 + x_2 + x_3 + \ldots + x_n) = E(x_1) + E(x_2) + E(x_3) + \ldots + E(x_n) = \mu + \mu + \mu + \ldots + \mu = n \times \mu$

無限基本單位數量之母體、樣本比例低($\frac{n}{N} \leq 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本和 s 抽樣分布的母體變異數$\sigma_s^2 = Var(s) = V(s) = V(x_1 + x_2 + x_3 + \ldots + x_n) = V(x_1) + V(x_2) + V(x_3) + \ldots + V(x_n) = \sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2 = n \times \sigma^2$

樣本和 s 抽樣分布的母體標準(偏)差 $\sigma_s = \sqrt{\sigma_s^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma$

有限基本單位數量之母體(抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本和 s 抽樣分布的母體變異數$\sigma_s^2 = n \times \sigma^2 \times \frac{N-n}{N-1}$

樣本和 s 抽樣分布的母體標準(偏)差 $\sigma_s = \sqrt{\sigma_s^2} = \sqrt{n \times \sigma^2 \times \frac{N-n}{N-1}} = \sigma \times \sqrt{n \times \frac{N-n}{N-1}}$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為有限母體校正因子(finite population correction factor)。當有限母體基本單位數量 N 很大，樣本數量 n 較小時，有限母體校正因子會趨近於 1。

範例 7.11　陸客到台灣觀光，每人每天購物消費金額 X 屬於常態分布(Normal distribution)，平均值 μ = NT\$ 8000 元，標準(偏)差 σ = NT\$ 2000 元。(A)隨機抽取一位大陸旅客當天購物消費金額超過 NT\$ 8500 元機率？(B)隨機抽取一團大陸旅客 30 人，平均每人當天購物消費超過 NT\$ 8500 元機率？(C)隨機抽取一團大陸旅客 20 人，全團當天購物消費總和不超過 NT\$ 150000 元機率？(D)隨機抽取一團大陸旅客 10 人，全團總和當天購物消費金額介在 NT\$ 60000 到 90000 元區間機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：大陸旅客到台灣觀光，隨機變數 X 代表每人每天購物消費金額的分布 $X \sim N(8000, 2000^2)$，母體平均值 μ = NT\$ 8000 元，母體標準(偏)差 σ = NT\$ 2000 元。

(A) $P(X > 8500) = P(\frac{x-\mu}{\sigma} > \frac{8500-8000}{2000}) = P(Z > 0.25) = 1 - P(Z \leq 0.25) = 1 - 0.5987$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.4013

(B) 樣本數量 $n = 30$：樣本平均值抽樣分布的標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{2000}{\sqrt{30}} = \frac{2000}{5.4772} = 365.1484$

$P(\bar{x} > 8500) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{8500-8000}{365.1484}) = P(Z > 1.3693) = 1 - P(Z \leq 1.3693) = 1 - 0.914548$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.0855

(C) 樣本數量 $n = 20$：樣本和抽樣分布的平均值$\bar{s} = E(s) = n \times \mu = 20 \times 8000 = 160000$。樣本和抽樣分布的標準(偏)差 $\sigma_s = \sqrt{\sigma_s^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma = \sqrt{20} \times 2000 = 4.4721 \times 2000 = 8944.2$

運用樣本和的運算方式 $P(s \leq 150000) = P(\frac{s-\bar{s}}{\sigma_s} \leq \frac{150000-160000}{8944.2}) = P(Z \leq -1.1180) = 0.1318$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

運用樣本平均值的運算方式 $P(s \leq 150000) = P(\bar{x} = \frac{s}{n} \leq \frac{150000}{20}) = P(\bar{x} \leq 7500) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \leq \frac{7500-8000}{\frac{2000}{\sqrt{20}}}) = P(Z \leq \frac{-500}{447.2136}) = P(Z \leq -1.1180) = 0.1318$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(D) 樣本數量 $n = 10$：樣本和抽樣分布的平均值$\bar{s} = E(s) = n \times \mu = 10 \times 8000 = 80000$。樣本和抽樣分布的標準(偏)差 $\sigma_s = \sqrt{\sigma_s^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma = \sqrt{10} \times 2000 = 3.1623 \times 2000 = 6324.6$

運用樣本和的運算方式 $P(60000 \leq s \leq 90000) = P(\frac{60000-80000}{6324.6} \leq \frac{s-\bar{s}}{\sigma_s} \leq \frac{90000-80000}{6324.6}) = P(-3.1623 \leq Z \leq 1.5811) = P(Z \leq 1.5811) - P(Z \leq -3.1623) = 0.9431 - 0.0008$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9423

運用樣本平均值的運算方式 $P(60000 \le s \le 90000) = P(\frac{60000}{10} \le \bar{x} = \frac{s}{n} \le \frac{90000}{10}) = P(6000 \le \bar{x} \le 9000) = P(\frac{6000-8000}{\frac{2000}{\sqrt{10}}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{9000-8000}{\frac{2000}{\sqrt{10}}}) = P(\frac{-2000}{632.4555} \le Z \le \frac{1000}{632.4555}) = P(-3.1623 \le Z \le 1.5811) = P(Z \le 1.5811) - P(Z \le -3.1623) = 0.9431 - 0.0008$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.9423$

答案：(A) 0.4013；(B) 0.0855；(C) 0.1318；(D) 0.9423

範例 7.12　大陸旅客到台灣觀光，每人每天購物消費金額 X 為常態分布(Normal distribution)，平均值 μ = NT\$ 8200 元，標準(偏)差 σ = NT\$ 1000 元。(A)隨機抽取一位大陸旅客當天購物消費金額超過 NT\$8500 元機率？(B)隨機抽取一團大陸客 300 人，平均每人當天購物消費少於 NT\$ 8200 元機率？(C)隨機抽取一團大陸客 30 人，平均每人當天購物消費超過 NT\$ 8500 元機率？(D)隨機抽取一團大陸客 40 人，平均每人當天購物消費等於 NT\$ 8200 元機率？(E)隨機抽取一團大陸旅客 20 人，全團當天購物消費總和不超過 NT\$ 150000 元機率？(F)隨機抽取一團大陸旅客 15 人，全團當天購物消費總和等於 NT\$ 120000 元機率？(G)隨機抽取一團大陸旅客 10 人，全團當天購物消費金額總和介在 NT\$ 75000 到 82000 元區間機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：大陸旅客到台灣觀光，隨機變數 X 代表每人每天購物消費金額的分布 $X \sim N(8200, 1000^2)$，母體平均值 μ = NT\$ 8200 元，母體標準(偏)差 σ = NT\$ 1000 元。

(A) $P(X > 8500) = P(\frac{X-\mu}{\sigma} > \frac{8500-8200}{1000}) = P(Z > 0.30) = 1 - P(Z \le 0.30) = 1 - 0.6179$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.3821$

(B) 樣本數量 $n = 300$，樣本平均值抽樣分布之母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{1000}{\sqrt{300}} = \frac{1000}{17.3205} = 57.7350$

$P(\bar{x} \le 8200) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{8200-8200}{57.7350}) = P(Z \le 0.000) = P(Z \le 0.000) = 0.5000$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(C) 樣本數量 $n = 30$，樣本平均值抽樣分布之母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{1000}{\sqrt{30}} = \frac{1000}{5.4772} = 182.5742$

$P(\bar{x} > 8500) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{8500-8200}{182.5742}) = P(Z > 1.643168) = 1 - P(Z \le 1.643168) = 1 - 0.9498$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0502$

(D) $P(\bar{x} = 8200) = 0.0000$，在連續機率分布中，特定一個點機率為 0.0000

(E) 樣本數量 $n = 20$，樣本和抽樣分布之平均值$\bar{s} = E(s) = n \times \mu = 20 \times 8200 = 164000$。樣本和抽樣分布之標準(偏)差 $\sigma_s = \sqrt{\sigma_S^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma = \sqrt{20} \times 1000 = 4.4721 \times 1000 = 4472.136$

運用樣本和的運算方式 $P(S \le 150000) = P(\frac{S-\bar{s}}{\sigma_s} \le \frac{150000-164000}{4472.136}) = P(Z \le -3.1305) = 0.0009$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

運用樣本平均值的運算方式 $P(S \le 150000) = P(\bar{x} = \frac{S}{n} \le \frac{150000}{20}) = P(\bar{x} \le 7500) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{7500-8200}{\frac{1000}{\sqrt{20}}}) = P(Z \le \frac{-700}{223.6068}) = P(Z \le -3.1305) = 0.0009$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(F) $P(S = 120000) = P(\bar{x} = \frac{S}{n} = \frac{120000}{15}) = P(\bar{x} = 8000) = 0.0000$，在連續機率分布中，特定一個點機率為 0.0000

(G) 樣本數量 $n = 10$，樣本和抽樣分布之平均值$\bar{s} = E(s) = n \times \mu = 10 \times 8200 = 82000$。樣本和抽樣分布之標準(偏)差 $\sigma_s = \sqrt{\sigma_S^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma = \sqrt{10} \times 1000 = 3.1623 \times 1000 = 3162.2777$

運用樣本和的運算方式 $P(75000 \le S \le 82000) = P(\frac{75000-82000}{3162.2777} \le \frac{S-\bar{s}}{\sigma_s} \le \frac{82000-82000}{3162.2777}) = P(-2.2136 \le Z \le 0) = P(Z \le 0) - P(Z \le -2.2136) = 0.5000 - 0.0134 = 0.4866 \fallingdotseq 0.4865$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

運用樣本平均值的運算方式 $P(75000 \le S \le 82000) = P(\frac{75000}{10} \le \bar{x} = \frac{S}{n} \le \frac{82000}{10}) = P(7500 \le \bar{x} \le 8200) = P(\frac{7500-8200}{\frac{1000}{\sqrt{10}}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{8200-8200}{\frac{1000}{\sqrt{10}}}) = P(\frac{-700}{316.2278} \le Z \le 0) = P(-2.2136 \le Z \le 0) = P(Z \le 0) - P(Z \le -2.2136)$ $= 0.5000 - 0.0134 = 0.4866$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

答案：(A) 0.3821；(B) 0.5000；(C) 0.0502；(D) 0.0000；(E) 0.0009；(F) 0.0000；(G) 0.4866

練習 7.15 奇遇飲料店今天製作 500 杯珍珠奶茶，其平均體積為 502.5 ml，標準(偏)差 3.5 ml。從今天製作的珍珠奶茶中，隨機抽出 100 杯為樣本。請估計(A)體積和在 50190 到 50200 ml 機率；(B)體積和超過 50300 ml 機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：運用樣本和的運算方式 $\bar{s} = E(s) = n \times \mu = 100 \times 502.5 = 50250$ ml。樣本和抽樣分布之標準(偏)差 $\sigma_s = \sqrt{n \times \sigma^2 \times \frac{N-n}{N-1}} = \sigma \times \sqrt{n \times \frac{N-n}{N-1}} = 3.5 \times \sqrt{100 \times \frac{500-100}{500-1}} = 3.5 \times 8.9532 = 31.3363$ ml

(A) $P(50190 \le S \le 50200) = P(\frac{50190-50250}{31.3363} \le \frac{S-\bar{s}}{\sigma_s} \le \frac{50200-50250}{31.3363}) = P(-1.9147 \le Z \le -1.59559) = P(Z \le -1.59559) - P(Z \le -1.9147) = 0.05529 - 0.02776$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0275$

(B) $P(S > 50300) = P(\frac{S-\bar{s}}{\sigma_s} > \frac{50300-50250}{31.3363}) = P(Z > \frac{50300-50250}{31.3363}) = P(Z > 1.595559) = 1 - P(Z \le 1.59559) = 1 - 0.9447$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0553$

運用樣本平均值的運算方式$\mu_{\bar{x}} = E(\bar{x}) = \mu = 502.5$ ml。樣本平均值抽樣分布之母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{3.5}{\sqrt{100}} \times \sqrt{\frac{500-100}{500-1}} = 0.313363$ ml

(A) $P(\frac{50190}{100} \le \bar{x} \le \frac{50200}{100}) = P(\frac{501.9-502.5}{0.313363} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{502-502.5}{0.313363}) = P(-1.9147 \le Z \le -1.59559) = P(Z \le -1.59559) - P(Z \le -1.9147) = 0.05529 - 0.02776$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0275$

(B) $P(\bar{x} > \frac{50300}{100}) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \ge \frac{503-502.5}{0.313363}) = P(Z > 1.595559) = 1 - P(Z \le 1.595559) = 1 - 0.9447$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0553$

答案：(A)$P(50190 \le S \le 50200) = 0.0275$；(B)$P(S > 50300) = 0.0553$

練習 7.16 台灣五星級飯店單人套房平均每日租金 NT$ 4250 元。單人套房每日租金分布不屬於常態分布，而是右偏分布。請估計(A)隨機抽取 45 間單人套房為樣本，樣本標準(偏)差為 NT$ 1235 元，平均每日租金低於 NT$ 3980 元機率；(B)隨機抽取 100 間單人套房為樣本，樣本標準(偏)差為 NT$ 2590 元，平均每日租金高於 NT$ 5000 元機率；(C)隨機抽取 100 間單人套房為樣本，樣本標準(偏)差為 NT$ 3590 元，平均每日租金介於 NT$ 2000 到 4560 元機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：隨機變數 X 代表台灣五星級飯店單人套房每日租金分布

(A)$P(\bar{x} \le 3980) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{3980-4250}{\frac{1235}{\sqrt{45}}}) = P(Z \le \frac{-270}{184.1029}) = P(Z \le -1.4666) = 0.0712$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(B)$P(\bar{x} > 5000) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{5000-4250}{\frac{2590}{\sqrt{100}}}) = P(Z > \frac{750}{259}) = P(Z > 2.8958) = 1 - P(Z \le 2.8958) = 1 - 0.9981$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.0019$

(C)$P(2000 \le \bar{x} \le 4560) = P(\frac{2000-4250}{\frac{3590}{\sqrt{100}}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{4560-4250}{\frac{3590}{\sqrt{100}}}) = P(\frac{-2250}{359} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{310}{359}) = P(-6.2674 \le Z \le 0.8635) = P(Z \le 0.8635) - P(Z \le -6.2674) = 0.8061 - 0.0000$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.8061$

答案：(A) 0.0712；(B) 0.0019；(C) 0.8061

練習 7.17 奇遇連鎖速食餐廳每位消費者消費金額分布屬於常態分布，其平均值為 NT$ 95.1 元，每位消費者消費金額的標準(偏)差 NT$ 15.4 元，現今隨機抽取 100 位消費者。試估算(A)100 位隨機樣本中，平均消費金額會在母體平均值±1 元區間內機率；(B)100 位隨機樣本中，平均消費金額會在母體平均值±2 元區間內機率；(C)100 位隨機樣本中，平均消費金額會在母體平均值±3 元區間內機率；(D)100 位隨機樣本中，全部消費金額總額會超過 NT$ 10000 元機率；(E)100 位隨機樣本中，全部消費金額總額會介在 NT$ 9200 到 9600 元區間機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：隨機變數 X 代表每位消費者消費金額分布。母體平均值 μ = NT$ 95.1 元，母體標準(偏)差 σ = NT$ 15.4 元，樣本數量 n = 100，樣本平均值抽樣分布之標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{15.4}{\sqrt{100}} = 1.54$。

(A) $P(95.1 - 1.0 \le \bar{x} \le 95.1 + 1.0) = P(94.1 \le \bar{x} \le 96.1) = P(\frac{94.1-95.1}{1.54} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{96.1-95.1}{1.54}) = P(-0.6494 \le Z \le 0.6494)$ $= P(Z \le 0.6494) - P(Z \le -0.6494) = 0.7419 - 0.2581$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.4839

(B) $P(95.1 - 2.0 \le \bar{x} \le 95.1 + 2.0) = P(93.1 \le \bar{x} \le 97.1) = P(\frac{93.1-95.1}{1.54} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{97.1-95.1}{1.54}) = P(-1.2987 \le Z \le 1.2987)$ $= P(Z \le 1.2987) - P(Z \le -1.2987) = 0.9030 - 0.0970$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.8060

(C) $P(95.1 - 3.0 \le \bar{x} \le 95.1 + 3.0) = P(92.1 \le \bar{x} \le 98.1) = P(\frac{92.1-95.1}{1.54} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{98.1-95.1}{1.54}) = P(-1.9481 \le Z \le 1.9481)$ $= P(Z \le 1.9481) - P(Z \le -1.9481) = 0.9743 - 0.0257$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9486

(D) 樣本和抽樣分布的平均值$\bar{s} = E(s) = n \times \mu = 100 \times 95.1 = 9510$。樣本和抽樣分布的母體標準(偏)差 $\sigma_s = \sqrt{\sigma_S^2} = \sqrt{n \times \sigma^2} = \sqrt{n} \times \sigma = \sqrt{100} \times 15.4 = 10 \times 15.4 = 154$

100 位隨機樣本中，全部消費金額總額會超過 NT$ 10000 元機率 $P(S > 10000) = P(\frac{S-\bar{s}}{\sigma_s} > \frac{10000-9510}{154}) = P(Z > 3.1818) = 1 - P(Z \le 3.1818) = 0.0007$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(E) 100 位隨機樣本中，全部消費金額總額會介在 NT$ 9200 到 9600 元區間機率 $P(92000 \le S \le 9600) = P(\frac{9200-9510}{154} \le \frac{S-\bar{s}}{\sigma_s} \le \frac{9600-9510}{154}) = P(-2.0130 \le Z \le 0.5844) = 0.7205 - 0.0221$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.6985

答案：(A) 0.4839；(B) 0.8060；(C) 0.9486；(D) 0.0007；(E) 0.6985

7.4.2.4 樣本平均值抽樣分布之應用

在應用樣本平均值的資料時，通常只有抽取一組的樣本，推估該樣本是否來自特定母體或來自特定母體機率，以推論該樣本是否來自特定母體。

範例 7.13 大陸旅客到台灣觀光，每人每天實際購物消費金額 X 屬於常態分布(Normal distribution)，平均值 μ = NT$ 8000 元，標準(偏)差 σ = NT$ 2000 元。(A)隨機抽取一團大陸旅客 30 人，統計平均每人當天購物消費金額 NT$ 9500 元(以上)，該團是否被導遊帶到黑店坑錢；(B)被坑錢機率？(答案有效位數四捨五入後取到小數點後第 4 位)

題解：大陸旅客到台灣觀光，隨機變數 X 代表每人每天購物消費金額的分布 $X \sim N(8000, 2000^2)$，母體平均值 μ = NT$ 8000 元，母體標準(偏)差 σ = NT$ 2000 元，樣本數量 n = 30。樣本平均值抽樣分布的母體標準(偏)差$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{2000}{\sqrt{30}} = \frac{2000}{5.4772} = 365.1484$。

平均每人當天購物消費金額，即樣本平均值$\bar{x}$ = 9500 元，高於母體平均值 μ = NT$ 8000 元，故必須計算其機率 $P(\bar{x} \ge 9500) = P(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \ge \frac{9500-8000}{365.1484}) = P(Z \ge 4.1079) = 1 - P(Z \le 4.1079) = 1 - 1$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.0000

平均值$\bar{x}$等於或大於 9500 元機率幾乎等於 0.0000，故不可能正常發生。極有可能被黑心導遊帶到黑心店，低價高售，大坑大陸旅客的錢，不合理。若被大陸旅客知情，台灣會被喧染成黑心之島，不利台灣觀光之永續發展。

答案：(A)平均值大於 9500 元機率幾乎等於 0.0000，極可能被帶到黑店；(B) 1.0000

練習 7.18　奇遇牌義式咖啡機製作咖啡的體積屬於常態分布，平均體積 100 ml，標準(偏)差 5 ml。現今製作 20 杯義式咖啡，測量其平均體積，請計算平均體積介在 98.2 和 102.5 ml 機率。(答案有效位數四捨五入後取到小數點後第 4 位)

題解：以隨機變數 X 代表咖啡飲料的體積，單位：ml。咖啡飲料體積 $X \sim N(100, 5^2)$

平均體積介在 98.2 和 102.5 ml 機率 $P(98.2 \le \bar{x} \le 102.5) = P(\frac{98.2-100}{\frac{5}{\sqrt{20}}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{102.5-100}{\frac{5}{\sqrt{20}}}) = P(-1.6100 \le Z \le 2.2361) = P(Z \le 2.2361) - P(Z \le -1.6100) = 0.9873 - 0.0537$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) $= 0.9336$

答案：機率 = 0.9336

7.5 樣本比率抽樣分布

在實務應用層面，當碰到間斷機率分布時，運用樣本比率(sample proportion)$\bar{p}$或$\hat{p}$ (讀音 p-hat)推估母體比率 p 數值。

母體中 N 個基本單位之隨機變數 X 屬於間斷機率分布，機率分布 $f(x)$，從母體中抽取 n 個基本單位為樣本，成功數量 x_1，成功樣本比例$\bar{p} = \frac{x_1}{n}$。若從母體 N 個基本單位抽出 n 個基本單位為樣本，會產生C_n^N種樣本的組合。

樣本比率$\bar{p}$之抽樣分布即是C_n^N種樣本組合的樣本比率$\bar{p}$數值機率分布。

- 7.5.1樣本比例之平均值與變異數
  - 無限基本單位數量
  - 有限基本單位數量
- 7.5.2樣本比例之抽樣分布
  - 7.5.2.1樣本數量與成功機率較高者
    - 無限基本單位數量
    - 有限基本單位數量
  - 7.5.2.2樣本數量與成功機率較低者
    - 無限基本單位數量
    - 有限基本單位數量

7.5.1 樣本比率分布之期望值與變異數

樣本比率$\bar{p}$抽樣分布的期望值 $E(\bar{p})$或母體平均值$\mu_{\bar{p}}$即等於原始資料間斷隨機變數 X 分布之母體比例 p。

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

無限基本單位數量之母體、樣本比例低($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \mathrm{Var}(\bar{p}) = V(\bar{p}) = \frac{p \times q}{n} = \frac{p \times (1-p)}{n}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p \times q}{n}} = \sqrt{\frac{p \times (1-p)}{n}}$

有限基本單位數量之母體(抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2$ = Var($\bar{p}$) = $V(\bar{p}) = \frac{p \times q}{n} \times \frac{N-n}{N-1} = \frac{p \times (1-p)}{n} \times \frac{N-n}{N-1}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p \times q}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{p \times q}{n}} \times \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{p \times (1-p)}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{p \times (1-p)}{n}} \times \sqrt{\frac{N-n}{N-1}}$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為有限母體校正因子(finite population correction factor)。

範例 7.14　小雯海產店過去 5 個月每月營業額與盈餘(Earnings, E)或虧損(Loss, L)的分布如下表：

月份	1	2	3	4	5
營業額	245000	255000	255000	265000	255000
盈餘或虧損	L	L	E	E	E

母體中盈餘(Earnings, E)或虧損(Loss, L)次數分布和機率分布

x	f	$f(x)$
E	3	0.6
L	2	0.4
合計	5	1.0

從過去 5 個月的數值，隨機抽取 3 個月的數值當樣本，請計算盈餘的比率抽樣分布之母體平均值$\mu_{\bar{p}}$及標準(偏)差$\sigma_{\bar{p}}$。

題解：盈餘樣本比例抽樣分布期望值 $E(\bar{p}) = \mu_{\bar{p}} = p = \frac{3}{5} = 0.6$。母體中 $N = 5$ 個基本單位，抽出 $n = 3$ 個當樣本，屬於有限母體。

盈餘樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2$ = Var($\bar{p}$) = $V(\bar{p}) = \frac{p \times (1-p)}{n} \times \frac{N-n}{N-1} = \frac{0.6 \times (1-0.6)}{3} \times \frac{5-3}{5-1} = 0.04$

盈餘樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p \times (1-p)}{n} \times \frac{N-n}{N-1}} = \sqrt{0.04} = 0.2000$

答案：盈餘樣本比例抽樣分布之平均值$\mu_{\bar{p}} = 0.6$；標準(偏)差$\sigma_{\bar{p}} = 0.2000$

7.5.2 樣本比率之抽樣分布型態

7.5.2.1 樣本數量與成功機率較高者($n \times p > 5$ 和 $n \times q > 5$)

中央極限定理『樣本數量 n 增加時，樣本比率$\bar{p}$之抽樣分布會接近於常態分布』。

樣本比率$\bar{p}$抽樣分布的期望值 $E(\bar{p})$或母體平均值$\mu_{\bar{p}}$即等於原始資料間斷隨機變數 X 分布的母體比例 p。

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

無限基本單位數量之母體、樣本比例低($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p \times q}{n} = \frac{p \times (1-p)}{n}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p \times q}{n}} = \sqrt{\frac{p \times (1-p)}{n}}$

樣本比率$\bar{p}$的抽樣分布趨近於常態分布，即樣本比率$\bar{p} \sim N(p, \frac{p \times q}{n})$。可以利用近似常態分布的方式運算樣本比例在特定條件下的出現機率。

有限基本單位數量之母體或樣本比例高($\frac{n}{N} > 0.05$) (抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2$ = Var($\bar{p}$) = V($\bar{p}$) = $\frac{p \times q}{n} \times \frac{N-n}{N-1} = \frac{p \times (1-p)}{n} \times \frac{N-n}{N-1}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}}$ = $\sqrt{\sigma_{\bar{p}}^2}$ = $\sqrt{\frac{p\times q}{n}\times\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times q}{n}}\times\sqrt{\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times(1-p)}{n}}\times\sqrt{\frac{N-n}{N-1}}$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為有限母體校正因子(finite population correction factor)。

若樣本數量 $n\leq 100$ 或更高數量時，利用常態分布的方法計算樣本比率$\bar{p}$在特定區間內機率時，必須進行**區間修正**。原始二項分布屬於間斷機率分布，僅在特定變量有機率的數值，欲使用連續機率分布中常態分布進行近似機率的運算時，因連續機率分布中必須運算變數在一定變量區間內才會有機率的數值，在利用近似常態分布的方式運算時，必須外擴 0.5 個單位，獲得的近似機率與利用二項分布運算的精準機率最接近。**建議在使用二項分布以近似常態分布方式計算區間機率時，都進行區間修正，所獲得機率數值與正確機率數值會比較接近。**

二項分布 $P(p_1 \leq \bar{p} \leq p_2) \approx$ 常態分布 $P[p_1-(\frac{1}{2}\times\frac{1}{n}) \leq \bar{p} \leq p_2+(\frac{1}{2}\times\frac{1}{n})]$ = 常態分布 $P(p_1-\frac{1}{2\times n} \leq \bar{p} \leq p_2+\frac{1}{2\times n})$

樣本比率$\bar{p}$的抽樣分布接近**常態分布**，即樣本比率$\bar{p}\sim N(p, \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1})$。可以利用近似常態分布的方式運算樣本比例在特定條件下的出現機率。

7.5.2.2 樣本數量與成功機率較低者($n\times p<5$ 或 $n\times q<5$)【選擇教材】

樣本比率$\bar{p}$抽樣分布的期望值 $E(\bar{p})$或母體平均值$\mu_{\bar{p}}$即等於原始資料間斷隨機變數 X 分布的母體比例 p。

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

無限基本單位數量之母體、樣本比例低($\frac{n}{N}\leq 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n} = \frac{p\times(1-p)}{n}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times q}{n}} = \sqrt{\frac{p\times(1-p)}{n}}$

樣本比率$\bar{p}$的抽樣分布屬於二項分布，即樣本比率$\bar{p}\sim$二項分布$(p, \frac{p\times q}{n})$

有限基本單位數量之母體或樣本比例高($\frac{n}{N}>0.05$) (抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n}\times\frac{N-n}{N-1} = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}$

樣本比率$\bar{p}$抽樣分布的母體標準(偏)差$\sigma_{\bar{p}}$ = $\sqrt{\sigma_{\bar{p}}^2}$ = $\sqrt{\frac{p\times q}{n}\times\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times q}{n}}\times\sqrt{\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}}$ = $\sqrt{\frac{p\times(1-p)}{n}}\times\sqrt{\frac{N-n}{N-1}}$

$\frac{N-n}{N-1}$和$\sqrt{\frac{N-n}{N-1}}$皆被稱為有限母體校正因子(finite population correction factor)。

樣本比率$\bar{p}$的抽樣分布屬於超幾何分布，即樣本比率$\bar{p}\sim$超幾何分布$(p, \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1})$

表 7-2 樣本比例$\bar{p}$之抽樣分布

母體屬性	樣本數量	條件	樣本比例的抽樣分布
二項分布 $E(X) = n\times p$	$n\times p>5$ 和 $n\times q>5$	無限基本單位數量之母體、樣本比例低($\frac{n}{N}\leq 0.05$)或 有限基本單位數量之母體(抽出的樣本放回)	$\bar{p}\sim N(p, \frac{p\times q}{n})$

母體屬性	樣本數量	條件	樣本比例的抽樣分布
$Var(X) = n \times p \times q$		有限基本單位數量之母體或樣本比例高($\frac{n}{N} > 0.05$)(抽出的樣本不放回)	$\bar{p} \sim N(p, \frac{p\times(1-p)}{n} \times \frac{N-n}{N-1})$
	$n \times p < 5$ 或 $n \times q < 5$	無限基本單位數量之母體、樣本比例低($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回)	$\bar{p}$~二項分布 B(n, p)
		有限基本單位數量之母體或樣本比例高($\frac{n}{N} > 0.05$)(抽出的樣本不放回)	$\bar{p}$~超幾何分布 Hypergeometric(N, r, n)

範例 7.15 高雄旗津欲興建跨港纜車，原經過市政府委託民間機構進行高雄市民意調查，宣稱高雄市市民的贊成百分比 65 % (反對率為 35 %)。今隨機訪談 200 位高雄市市民，贊成百分比在 60~70 %機率為多少？(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體為高雄市民，N 值非常大，$\frac{n}{N} \le 0.05$。故，樣本比率$\bar{p}$的抽樣分布趨近於**常態分布**。贊成機率 $p = 0.65$，樣本數量 $n = 200$。

贊成樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times q}{n} = \frac{p\times(1-p)}{n} = \frac{0.65\times(1-0.65)}{200} = \frac{0.2275}{200} = 0.0011375$

贊成樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times q}{n}} = \sqrt{\frac{p\times(1-p)}{n}} = \sqrt{0.0011375} = 0.0337$

贊成百分比在 60~70 %機率 $P(0.6 \le \bar{p} \le 0.7) \approx$ 常態分布 $P(\frac{0.6-0.65}{0.0337} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.7-0.65}{0.0337}) = P(-1.4837 \le Z \le 1.4837) = P(Z \le 1.4837) - P(Z \le -1.4837) = 0.9311 - 0.0689$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.8622 (利用常態分布運算未進行區間修正的近似答案)

二項分布以趨近於常態分布的運算過程，進行**區間修正**：二項分布 $P(0.6 \le \bar{p} \le 0.7) \approx$ 常態分布 $P(0.6 - \frac{1}{2} \times \frac{1}{200} \le \bar{p} \le 0.7 + \frac{1}{2} \times \frac{1}{200})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(0.5975 \le \bar{p} \le 0.7025) = P(\frac{0.5975-0.65}{0.0337} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.7025-0.65}{0.0337}) = P(-1.5566 \le Z \le 1.5566) = P(Z \le 1.5566) - P(Z \le -1.5566) = 0.9402 - 0.0598$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.8804 (利用常態分布運算的近似答案)

直接以二項分布的模式運算：**在 200 位高雄市民的樣本中，贊成 60 %代表：200 × 60 % = 120 人贊成；贊成 70%代表：200 × 70 % = 140 人贊成**。【若依據百分比運算出來的人數不是整數時，下限人數則是無條件進位成整數，上限人數則是無條件移除小數，例如：在 202 位高雄市民的樣本中，贊成 60 %代表：202 × 60 % = 121.2，則要運算 122 人(含)以上贊成；贊成 70%代表：202 × 70 % = 141.4，運算 141 人(含)以下贊成。】隨機變數 X 代表贊成人數。

贊成百分比在 60~70 %機率 $P(120 \le X \le 140) = P(X \le 140) - P(X \le 119) = 0.9416 - 0.0609$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.8807 (精準答案)

答案：贊成百分比在 60~70 %機率 0.8807

樣本數量	未區間修正	區間修正	二項分布	未修正與二項差距	修正與二項差距
50	0.54146	0.62627	0.62617	0.08471	0.00010
100	0.70549	0.75114	0.75138	0.04589	0.00024
150	0.80082	0.82915	0.82946	0.02864	0.00031
200	0.86179	0.88044	0.88074	0.01895	0.00030
250	0.90258	0.91525	0.91552	0.01294	0.00027
300	0.93058	0.93937	0.93960	0.00902	0.00022
350	0.95014	0.95632	0.95651	0.00637	0.00018
400	0.96397	0.96836	0.96851	0.00454	0.00015
450	0.97383	0.97698	0.97710	0.00327	0.00012

樣本數量	未區間修正	區間修正	二項分布	未修正與二項差距	修正與二項差距
500	0.98092	0.98319	0.98329	0.00236	0.00009
550	0.98605	0.98769	0.98776	0.00172	0.00007
600	0.98976	0.99096	0.99102	0.00125	0.00005
650	0.99247	0.99335	0.99339	0.00092	0.00004
700	0.99445	0.99509	0.99513	0.00067	0.00003
750	0.99591	0.99638	0.99640	0.00049	0.00002
800	0.99697	0.99732	0.99734	0.00036	0.00002
850	0.99776	0.99801	0.99803	0.00027	0.00001
900	0.99834	0.99853	0.99854	0.00020	0.00001

當樣本基本單位數量 n 愈多時，修正的係數($\frac{1}{2}\times\frac{1}{n}$)數值就會愈低，區間修正與未修正者，兩者差異會愈來愈小。

練習 7.19　高雄市失業率為 6.0 %，若從高雄市中隨機抽出 200 位居民調查，請計算樣本失業率至少 5 %機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體比例 $p = 0.06$，樣本數量 $n = 200$。

樣本 $n = 200$ 失業比率$\bar{p} = 0.05$，對應的失業人數 $= n \times \bar{p} = 200 \times 0.05 = 10$ 人。【若依據百分比運算出來的人數不是整數時，人數則是無條件進位成整數，例如：在 202 位高雄市居民的樣本中，樣本失業率至少 5 %代表：202 × 5 % = 10.1，則要運算 11 人(含)以上失業。】使用隨機變數 X 代表在該樣本數量下失業人數。

二項分布運算：樣本失業率至少 5 %機率 $P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.2343$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.7657 (精準答案)

近似常態分布運算：二項分布 $P(\bar{p} \geq 0.05) \approx$ 常態分布 $P(\bar{p} \geq 0.05 - \frac{1}{2}\times\frac{1}{200})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \geq 0.0475)$ = 常態分布 $P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \geq \frac{0.0475-0.06}{\sqrt{\frac{0.06\times0.94}{200}}}) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \geq \frac{0.0475-0.06}{0.0168}) = P(Z \geq -0.7444) = 1 - P(Z \leq -0.7444) = 1 - 0.2283$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.7717 (近似答案)

答案：樣本失業率至少 5 %機率 $P(\bar{p} \geq 0.05) = 0.7657$

範例 7.16　高雄旗津欲興建跨港纜車，原經過市政府委託民間機構進行高雄市民意調查，宣稱高雄市市民的贊成百分比 65 % (反對率為 35 %)。今隨機訪談 20 位高雄市市民，贊成百分比在 60~70 %機率為多少？(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體為高雄市民，N 值非常大，$\frac{n}{N} \leq 0.05$。故，樣本比率$\bar{p}$的抽樣分布屬於二項分布。贊成機率 $p = 0.65$，樣本數量 $n = 20$，隨機變數 X 代表贊成人數。

20 位高雄市市民贊成 60 %人數：20 × 60 % = 12 人贊成；贊成 70 %人數：20 × 70 % = 14 人贊成。若依據百分比運算出來的人數不是整數時，下限人數則是無條件進位成整數，上限人數則是無條件移除小數，例如：在 21 位高雄市民的樣本中，贊成 60 %代表：21 × 60 % = 12.6，則要運算 13 人(含)以上贊成；贊成 70%代表：21 × 70 % = 14.7，運算 14 人(含)以下贊成。

贊成百分比在 60~70 %機率 $P(12 \leq X \leq 14) = P(X \leq 14) - P(X \leq 11) = 0.7546 - 0.2376$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.5170

答案：贊成百分比在 60~70 %機率 0.5170

範例 7.17 高雄旗津欲興建跨港纜車，原經過市政府委託民間機構進行 300 位觀光相關學者民意調查，宣稱觀光相關學者的贊成百分比 65 % (反對率為 35 %)。今隨機訪談 50 位觀光相關學者，贊成百分比在 60~70 機率為多少？(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體為 300 位觀光相關學者，N 值不大，$\frac{n}{N} = \frac{50}{300} = 0.16667 > 0.05$。故，贊成樣本比率$\bar{p}$的抽樣分布接近**常態分布**。贊成機率 $p = 0.65$，樣本數量 $n = 50$(樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

贊成樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n} \times \frac{N-n}{N-1} = \frac{0.65\times(1-0.65)}{50} \times \frac{300-50}{300-1} = \frac{0.2275}{50} \times \frac{250}{299} = 0.003804$

贊成樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n} \times \frac{N-n}{N-1}} = \sqrt{0.003804} = 0.061679$

近似常態分布方式運算：二項分布 $P(0.60 \le \bar{p} \le 0.70) \approx$ 常態分布 $P(0.60 - \frac{1}{2\times50} \le \bar{p} \le 0.70 + \frac{1}{2\times50})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(0.59 \le \bar{p} \le 0.71) = P(\frac{0.59-0.65}{0.061679} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.71-0.65}{0.061679}) = P(-0.97277 \le Z \le 0.97277) = P(Z \le 0.97277) - P(Z \le -0.97277) = 0.834667 - 0.165333$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.6693 (近似答案)

二項分布的模式運算：在 300 位觀光相關學者母體資料中，隨機抽出 50 位學者，贊成 60%代表：50 × 60 % = 30 人贊成；贊成 70%代表：50 × 70 % = 35 人贊成。隨機變數 X 代表贊成人數。

贊成百分比在 60~70 %機率 $P(30 \le X \le 35) = P(X \le 35) - P(X \le 29) = 0.8122 - 0.1861$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.6262 (精準答案)

顯示在母體總數不大，樣本數量也不大的情況下，若二項分布已趨近於常態分布的方式運算，也進行區間修正，還是會有些許的誤差量。

答案：二項分布 $P(0.60 \le \bar{p} \le 0.70) = 0.6262$

練習 7.20 若母體比率為 0.65，請分別計算在樣本數量 n 等於 100、500、1000 和 1500 的樣本比率之標準(偏)差？當樣本數增加時，樣本比率之標準(偏)差有何變異？

練習 7.21 高雄旗津欲興建跨港纜車，原經過市政府委託民間機構進行 200 位觀光相關學者民意調查，宣稱觀光相關學者的贊成比率 0.55 (反對率為 0.45)。試評估(A)今隨機訪談 50 位觀光相關學者，贊成比率在 0.60~0.70 機率；(B)今隨機訪談 20 位觀光相關學者，贊成比率在 0.60 以下機率；(C)今隨機訪談 30 位觀光相關學者，贊成比率在 0.70 以上機率；(D)今隨機訪談 15 位觀光相關學者，贊成比率為 0.70 機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：母體為 200 位觀光相關學者，N 值不大，$\frac{n}{N} = \frac{50}{200} = 0.25 > 0.05$。故，樣本比率$\bar{p}$的抽樣分布接近常態分布。

(A)贊成機率 $p = 0.55$，樣本數量 $n = 50$ (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n} \times \frac{N-n}{N-1} = \frac{0.55\times(1-0.55)}{50} \times \frac{200-50}{200-1} = \frac{0.2475}{50} \times \frac{150}{199} = 0.0037311$

樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n} \times \frac{N-n}{N-1}} = \sqrt{0.0037311} = 0.0610827$

贊成比率在 0.60~0.70 機率：二項分布 $P(0.60 \le \bar{p} \le 0.70) \approx$ 常態分布 $P(0.60 - \frac{1}{2\times50} \le \bar{p} \le 0.70 + \frac{1}{2\times50})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(0.59 \le \bar{p} \le 0.71) = P(\frac{0.59-0.55}{0.0610827} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.71-0.55}{0.0610827}) = P(0.6548 \le Z \le 2.6194) = P(Z \le 2.6194) - P(Z \le 0.6548) = 0.9956 - 0.7437$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.2519

反對機率 $p = 0.45$，樣本數量 $n = 50$ (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

反對樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n} \times \frac{N-n}{N-1} = \frac{0.45\times(1-0.45)}{50} \times \frac{200-50}{200-1} = \frac{0.2475}{50} \times \frac{150}{199} = 0.0037311$

反對樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{0.0037311} = 0.0610827$

反對比率在 0.30~0.40 機率：二項分布 $P(0.30 \le \bar{p} \le 0.40) \approx$ 常態分布 $P(0.30 - \frac{1}{2\times50} \le \bar{p} \le 0.40 + \frac{1}{2\times50})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(0.29 \le \bar{p} \le 0.41) = P(\frac{0.29-0.45}{0.0610827} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.41-0.45}{0.0610827}) = P(-2.6194 \le Z \le -0.6548) = P(Z \le -0.6548) - P(Z \le -2.6194) = 0.2563 - 0.0044$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.2519

(B)贊成機率 p = 0.55，樣本數量 n = 20 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)

贊成樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1} = \frac{0.55\times(1-0.55)}{20}\times\frac{200-20}{200-1} = \frac{0.2475}{20}\times\frac{180}{199} = 0.011193$

贊成樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{0.011193} = 0.105799$

贊成比率在 0.60 以下機率：二項分布 $P(\bar{p} \le 0.60) \approx$ 常態分布 $P(\bar{p} \le 0.60 + \frac{1}{2\times20})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \le 0.625) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.625-0.55}{0.105799}) = P(Z \le 0.70889) = 0.760804$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

反對機率 p = 0.45，樣本數量 n = 20 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

反對樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1} = \frac{0.45\times(1-0.45)}{20}\times\frac{200-20}{200-1} = \frac{0.2475}{20}\times\frac{180}{199} = 0.011193$

反對樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{0.011193} = 0.105799$

反對比率在 0.40 以上機率：二項分布 $P(\bar{p} \ge 0.40) \approx$ 常態分布 $P(\bar{p} \ge 0.40 - \frac{1}{2\times20})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \ge 0.375) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.375-0.45}{0.105799}) = P(Z \ge -0.70889) = 1 - P(Z \le -0.70889) = 1 - 0.239196$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.760804

(C)贊成機率 p = 0.55，樣本數量 n = 30 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)

贊成樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1} = \frac{0.55\times(1-0.55)}{30}\times\frac{200-30}{200-1} = \frac{0.2475}{30}\times\frac{170}{199} = 0.007048$

贊成樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{0.007048} = 0.083951$

贊成比率在 0.70 以上機率：二項分布 $P(\bar{p} \ge 0.70) \approx$ 常態分布 $P(\bar{p} \ge 0.70 - \frac{1}{2\times30})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \ge 0.683333) = 1 - P(\bar{p} \le 0.68333) = 1 - P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.683333-0.55}{0.083951}) = 1 - P(Z \le 1.58823) = 1 - 0.9439$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.0561

反對機率 p = 0.45，樣本數量 n = 30 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

反對樣本比例抽樣分布的母體變異數$\sigma_{\bar{p}}^2 = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1} = \frac{0.45\times(1-0.45)}{30}\times\frac{200-30}{200-1} = \frac{0.2475}{30}\times\frac{170}{199} = 0.007048$

反對樣本比例抽樣分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{0.007048} = 0.083951$

反對比率在 0.30 以下機率：二項分布 $P(\bar{p} \le 0.30) \approx$ 常態分布 $P(\bar{p} \le 0.30 + \frac{1}{2\times30})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \le 0.316667) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.316667-0.45}{0.083951}) = P(Z \le -1.58823) = 0.0561$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得)

(D)贊成機率 p = 0.55，樣本數量 n = 15 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)

今隨機訪談 15 位觀光相關學者，贊成比率為 0.70 機率，15 × 0.70 = 10.5 位觀光相關學者贊成機率是 0，因為在二項分布中，非整數對應機率皆是 0。二項分布 $P(\bar{p} = 0.70) = 0.0000$。

反對機率 p = 0.45，樣本數量 n = 15 (樣本比例抽樣分布的變異數和標準差需要進行有限母體校正)。

今隨機訪談 15 位觀光相關學者，反對比率為 0.30 機率，15 × 0.30 = 4.5 位觀光相關學者贊成機率是 0，因為在二項分布中，非整數對應機率皆是 0。二項分布 $P(\bar{p} = 0.30) = 0.0000$。

答案：(A)二項分布 $P(0.60 \le \bar{p} \le 0.70) = 0.2519$；(B)二項分布 $P(\bar{p} \le 0.60) = 0.7608$；(C)二項分布 $P(\bar{p} \ge 0.70) = 0.0561$；(D)二項分布 $P(\bar{p} = 0.70) = 0.0000$

練習 7.22 全球經濟不景氣，消費慾望低，失業率高昇，高雄市現階段有 6 %失業率，現階段隨機抽樣選擇 500 位市民，進行失業率調查。(A)若母體失業比率 $p = 0.06$，請敘述 $n = 500$ 樣本失業比率$\bar{p}$機率分布(利用平均值和標準(偏)差表示)；(B) $n = 500$ 樣本失業比率$\bar{p}$超過 0.08 機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：

(A)母體高雄市民數量巨大，樣本數量 $n = 500$，失業率為 $p = 0.06$。樣本失業比率$\bar{p}$的抽樣分布接近常態分布。$E(\bar{p}) = \mu_{\bar{p}} = p = 0.06$。

失業樣本比例分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\frac{p\times(1-p)}{n}} = \sqrt{\frac{0.06\times(1-0.06)}{500}} = \sqrt{0.000113} = 0.01062$

(B)樣本 $n = 500$ 失業比率$\bar{p} = 0.08$，對應的失業人數 $= n \times \bar{p} = 500 \times 0.08 = 40$ 人。使用隨機變數 X 代表在該樣本數量下失業人數。

$n = 500$ 樣本失業比率$\bar{p}$超過 0.08 機率 $P(X \ge 41) = 1 - P(X \le 40) = 1 - 0.9719$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.0281 (精準答案)

二項分布 $P(\bar{p} \ge 0.08) \approx$ 常態分布 $P(\bar{p} \ge 0.08 - \frac{1}{2} \times \frac{1}{500})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \ge 0.0790)$ = 常態分布 $P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.0790-0.06}{\sqrt{\frac{0.06\times0.94}{500}}}) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.0790-0.06}{0.01062}) = P(Z \ge 1.78895) = 1 - P(Z \le 1.78895) = 1 - 0.9632$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.0368 (使用常態分布運算近似答案)

答案：(A)平均值 = 0.0600、標準(偏)差 = 0.0106；(B)機率 = 0.0368

練習 7.23 純純廠牌咖啡機有 3.0 %的不良率，今隨機抽樣該廠牌咖啡機 90 台，請估算(A)至少有 4.2 %不良品機率 $P(\bar{p} \ge 0.042)$；(B)至多有 2.0 %不良品機率 $P(\bar{p} \le 0.02)$？(答案有效位數四捨五入取到小數點後第 4 位)

題解：咖啡機數量生產眾多，樣本數量 $n = 90$(需要進行區間修正)，不良機率 $p = 0.03$。故，樣本不良比率$\bar{p}$的抽樣分布接近常態分布。樣本比例的期望值 $E(\bar{p}) = \mu_{\bar{p}} = p = 0.03$

不良品樣本比例分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\frac{p\times(1-p)}{n}} = \sqrt{\frac{0.03\times(1-0.03)}{90}} = \sqrt{0.000323} = 0.017981$

(A)隨機抽樣該廠牌咖啡機 90 台，有 4.2 %不良品，對應的不良品數量 90 × 4.2 % = 3.78 台。使用隨機變數 X 代表在該樣本數量下不良品數量。

至少有 4.2 %不良品機率 $P(X \ge 4) = 1 - P(X \le 3) = 1 - 0.7151$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.2849 (精準答案)

二項分布 $P(\bar{p} \ge 0.042) \approx$ 常態分布 $P(\bar{p} \ge 0.042 - \frac{1}{2} \times \frac{1}{90})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \ge 0.036444)$ = 常態分布 $P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.036444-0.03}{\sqrt{\frac{0.03\times0.97}{90}}}) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.036444-0.03}{0.017981}) = P(Z \ge 0.3584) = 1 - P(Z \le 0.3584) = 1 - 0.63997$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.3600 (使用常態分布運算近似答案)

(B)隨機抽樣該廠牌咖啡機 90 台，有 2.0 %不良品，對應的不良品數量 90 × 2.0 % = 1.8 台。使用隨機變數 X 代表在該樣本數量下不良品數量。

至多有 2.0 %不良品機率 $P(X \le 1) = 0.2440$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) (精準答案)

二項分布 $P(\bar{p} \le 0.02) \approx$ 常態分布 $P(\bar{p} \le 0.02 + \frac{1}{2} \times \frac{1}{90})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \le 0.025556)$ = 常態分布 $P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.025556-0.03}{\sqrt{\frac{0.03\times0.97}{90}}}) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.025556-0.03}{0.017981}) = P(Z \le -0.24717) = 0.4024$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) (使用常態分布運算近似答案)

答案：(A)至少有 4.2 %不良品機率 $P(\bar{p} \ge 0.042) = 0.3600$；(B)至多有 2.0 %不良品機率 $P(\bar{p} \le 0.02) = 0.4024$

練習 7.24　大純旅館出租客房中 30.0 %來自於團體性客人，今隨機抽樣 98 間客房，請估算(A)團體性客人比例介於 0.2 到 0.4 之間機率；(B)團體性客人比例超過 0.38 機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：旅館出租房客數量眾多，樣本數量 $n = 98$(需要進行區間修正)，團體客人機率 $p = 0.30$。故，樣本團體客人比率$\bar{p}$的抽樣分布接近常態分布。團體客樣本比例分布的期望值 $E(\bar{p}) = \mu_{\bar{p}} = p = 0.30$。

團體客樣本比例分布的母體標準(偏)差$\sigma_{\bar{p}} = \sqrt{\frac{p\times(1-p)}{n}} = \sqrt{\frac{0.30\times(1-0.30)}{98}} = \sqrt{0.002143} = 0.046291$

(A)抽樣 98 間客房，團體客人比例介於 0.2 到 0.4，對應的客房數量 98 × 0.2 = 19.6 間房；98 × 0.4 = 39.2 間客房。使用隨機變數 X 代表團體性客人住房數量。

團體性客人住房數量 20~39 間機率 $P(20 \le X \le 39) = P(X \le 39) - P(X \le 19) = 0.9853 - 0.0122$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.9732 (精準答案)

二項分布 $P(0.20 \le \bar{p} \le 0.40) \approx$ 常態分布 $P(0.20 - \frac{1}{2\times98} \le \bar{p} \le 0.40 + \frac{1}{2\times98})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(0.194898 \le \bar{p} \le 0.405102) = P(\frac{0.194898-0.30}{0.046291} \le \frac{\bar{p}-p}{\sigma_{\bar{p}}} \le \frac{0.405102-0.30}{0.046291}) = P(-2.270464 \le Z \le 2.270464) = P(Z \le 2.270464) - P(Z \le -2.270464) = 0.98841 - 0.01159$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.9768 (使用常態分布運算近似答案)

(B)抽樣 98 間客房，團體性客人比例超過 0.38，對應的客房數量 98 × 0.38 = 37.24 間房。使用隨機變數 X 代表團體性客人住房數量。

團體性客人比例超過 0.38 機率 $P(X \ge 38) = 1 - P(X \le 37) = 1 - 0.9607$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.0393 (精準答案)

二項分布 $P(\bar{p} \ge 0.38) \approx$ 常態分布 $P(\bar{p} \ge 0.38 - \frac{1}{2} \times \frac{1}{98})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = 常態分布 $P(\bar{p} \ge 0.374898)$ = 常態分布 $P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.374898-0.30}{\sqrt{\frac{0.30\times0.70}{98}}}) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \ge \frac{0.374898-0.30}{0.046291}) = P(Z \ge 1.617981) = 1 - P(Z \le 1.617981) = 1 - 0.947167$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.052833 (使用常態分布運算近似答案)

答案：(A) $P(0.20 \le \bar{p} \le 0.40) = 0.9768$；(B) $P(\bar{p} \ge 0.38) = 0.0528$

練習 7.25　小琦咖啡館有自動咖啡機，製作相關咖啡飲料上萬杯，每日抽驗 100 杯，若達到 8 %的不良率，即停機檢修。若今天自動咖啡機的實際不良率為 6 %，請估算經抽驗不合於標準需要停機檢修機率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：合格與不合格屬於二項分布，不良機率 $p = 0.06$，樣本數量 $n = 100 > 30$，$n \times p = 6 > 5$，故，樣本比率的抽樣分布$\bar{p}$~$N(p, \frac{p\times q}{n})$。

樣本比例分布之母體變異數$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n} = \frac{p\times(1-p)}{n} = \frac{0.06\times(1-0.06)}{100} = \frac{0.0564}{100} = 0.000564$

樣本比例分布之母體標準差$\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2} = \sqrt{\frac{p\times q}{n}} = \sqrt{\frac{p\times(1-p)}{n}} = \sqrt{0.000564} = 0.0237$

未進行區間校正 $P(\bar{p} \geq 0.08) \approx P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \geq \frac{0.08-0.06}{0.0237}) = P(Z \geq 0.8422) = 1 - P(Z \leq 0.8422) = 1 - 0.8001$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.1999

區間校正：二項分布 $P(\bar{p} \geq 0.08) \approx$ 常態分布 $P(\bar{p} \geq 0.08 - \frac{1}{2} \times \frac{1}{100})$(樣本數量不是非常大的情況下，原始二項分布利用近似常態分布的方式計算需要進行區間修正) = $P(\bar{p} \geq 0.08 - \frac{1}{2} \times \frac{1}{100}) = P(\bar{p} \geq 0.075) = P(\frac{\bar{p}-p}{\sigma_{\bar{p}}} \geq \frac{0.075-0.06}{0.0237}) = P(Z \geq 0.6316) = 1 - P(Z \leq 0.6316) = 1 - 0.7362$(利用 Excel 軟體 NORM.S.DIST 函數查詢獲得) = 0.2638 (使用常態分布運算近似答案)

每日抽驗 100 杯，8 %的不良率，對應的不良品杯數 = 100 × 8 % = 8 杯。使用隨機變數 X 代表不良品杯數。

至少有 8 %的不良率，需要停機檢修機率 $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.7483$(使用 Excel 軟體中 BINOM.DIST 函數查詢獲得) = 0.2517 (精準答案)

答案：$P(\bar{p} \geq 0.08) = 0.2638$

練習 7.26　請將下列統計值連到其所對應參數。連連看。

	統計值	參數
(a)	S	μ
(b)	S^2	$\sigma_{\bar{x}}$
(c)	$\bar{x}$	P
(d)	$\bar{p}$	σ
(e)	$S_{\bar{x}}$	σ^2

練習 7.27　從平均值 μ 的母體中抽出三批樣本。樣本 A 樣本數量 $n = 20$，樣本 B 樣本數量 $n = 50$，樣本 C 樣本數量 $n = 100$。請問哪一個樣本平均值會比較接近於母體平均值 μ？為什麼？

練習 7.28　是非題(×)參數是透過樣本資料計算獲得的數值，可以呈現此樣本特定特徵。

練習 7.29　假設一個袋子裡有五粒球，球的外表都一樣，球的編號為 0, 1, 1, 1, 2；假設某甲隨機分別抽出兩顆球，抽球方式是允許放回方式(with replacement)。假設 X 與 X 分別表示某甲第一次抽的球號與第二次抽到的球號。(A)請計算出 $\bar{X} = \frac{X_1+X_2}{2}$之抽樣分配？(B)請計算出 S^2 之抽樣分配？(提示 $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$)(99 年普考統計學概要)

誤差異限(margin of error)

討論議題

1.師生在非同步教學討論議題：簡單隨機抽樣應用

已經完成簡單隨機抽樣的學習，簡單隨機抽樣必須具備【無限母體數量的簡單隨機抽樣必須確定樣本中每一個基本單位皆來自相同的母體，和樣本中每一個基本單位皆獨立抽出且機率相等。】的特性。

第一回合請於 D+3 日中午 1200 點以前從「議題討論」區【張貼】標題：「夜市消費者隨機抽樣」，本文：請選擇一個您最熟悉的夜市，詮釋如何運用(類似)簡單隨機抽樣(運用工具：彩球或彩券)獲得有代表性的夜市消費者的樣本，對象、時間分布、地點分布(具體街口、出入口)請具體明確的陳述(20 個字以上)。期望可以透過學習運用情境的交流，相互激勵，提升學習效益。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，詳細檢視其他同學的張貼內容。第二回合【張貼】標題：「最佳詮釋」，本文：自己靜下心來，集思廣益，思考一下，哪一位同學(自己除外)詮釋得最好，其理由(10 個字以上)。透過同學間分享討論，可以提升學習效益，加油！第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

2.學習者間同步討論議題：評量成績好的關鍵因素

第一回合請於 D 日早上 1100 點以前從「議題討論」區【張貼】標題：「關鍵因素」，本文：經歷將近一個學期各式各樣的上課練習和平常考，每位同學累積了相當多的應考經驗，請分享一下考高分的關鍵因素(20 個字以上詮釋)。期望可以透過應考經驗的分享學習與相互交流，相互激勵，提升學習效益。

待有 35 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，檢視其他同學的回應內容。第二回合【張貼】標題：「最佳詮釋」，本文：自己靜下心來，集思廣益，思考一下，哪一位同學詮釋得最好，其理由(15 個字以上)。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

3.學習者間同步討論議題：學習資源運用效益

第一回合請於 D 日早上 0910 點以前從「議題討論」區【張貼】標題：「資源效益」，本文：經歷上課前預習書本教材、數位影音教材、前一個年度上課錄影、牛刀小試，上課中同步講解和議題討論，上課後提供上課錄影、上課練習、平常考、考畢閱覽、來賓考試等學習資源，哪一種或兩種資源對你的學習最有效益？理由(15 個字以上詮釋)。期望可以透過教學資源的使用經驗分享與相互交流，相互激勵，提升學習效益。

待有 40 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，檢視其他同學的回應內容。即可開始張貼第二回合【張貼】標題：「最大領悟」，本文：自己靜下心來，集思廣益，思考一下，在學習資源的運用效益分享內容中，論述給自己最大領悟(10 個字以上詮釋)。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

重點整理

Excel 函數彙整

Excel 函數	統計功能	輸入資料	輸出資料
RAND	隨機樣本	無	介於 0 和 1 之間的隨機數值
RANDBETWEEN	隨機樣本	起始值和結束值	前述兩數值之間的隨機數值

在隨機抽樣過程中，抽取母體中每一個基本單位被抽中機率皆獨立且相等。藉由簡單隨機抽樣所獲得的樣本稱為簡單隨機樣本(simple random sample)或隨機樣本。

從母體中隨機抽出樣本，由樣本獲得的統計值(statistic)以作為(提供)母體參數(parameter)之估計值(estimate)。

點估計特性(Properties of point estimators)：一致性(Consistency)、無偏性(Unbiasedness)、有效性(Efficiency)、充分性(Sufficiency)

樣本平均值$\bar{x}$與母體平均值 μ 的**差異**即可視為抽樣誤差(sampling error) = $|\bar{x} - \mu|$。

C_n^N種樣本組合之樣本平均值$\bar{x}$的平均值，利用$\mu_{\bar{x}}$或 $E(\bar{x})$符號表示，即$\mu_{\bar{x}} = E(\bar{x}) = \mu$。

$$\mu_{\bar{x}} = E(\bar{x}) = E(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n} \times E(\sum_{i=1}^{n} x_i) = \frac{1}{n} \times E(x_1 + x_2 + x_3 + \ldots + x_n) = \frac{1}{n} \times [E(x_1) + E(x_2) + E(x_3) + \ldots + E(x_n)] = \frac{1}{n} \times (\mu + \mu + \mu + \ldots + \mu) = \frac{1}{n} \times n \times \mu = \mu$$

C_n^N種樣本組合之樣本平均值$\bar{x}$的變異數，利用$\sigma_{\bar{x}}^2$、$Var(\bar{x})$或 $V(\bar{x})$表示；C_n^N種樣本組合之樣本平均值$\bar{x}$的標準(偏)差(standard deviation of the sampling distribution of the sample mean)，利用$\sigma_{\bar{x}}$表示。

$$\sigma_{\bar{x}}^2 = Var(\bar{x}) = V(\bar{x}) = V(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n^2} \times V(\sum_{i=1}^{n} x_i) = \frac{1}{n^2} \times V(x_1 + x_2 + x_3 + \ldots + x_n)$$

$$= \frac{1}{n^2} \times [V(x_1) + V(x_2) + V(x_3) + \ldots + V(x_n)] = \frac{1}{n^2} \times (\sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

無限基本單位數量 N 之母體、($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，供下一個樣本抽取)：樣本平均值$\bar{x}$的變異數$\sigma_{\bar{x}}^2$與標準(偏)差$\sigma_{\bar{x}}$

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

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

有限基本單位數量 N 之母體：樣本平均值$\bar{x}$的變異數$\sigma_{\bar{x}}^2$與標準(偏)差$\sigma_{\bar{x}}$

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

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

中央極限定理(Central limit theorem, CLT)：母體平均值為 μ，標準(偏)差為 σ(μ 及 σ 皆為有限數值)的母體中，當樣本數量愈大($n \ge 30$)時，樣本平均值$\bar{x}$的分布愈接近(趨近、近似)常態分布，且向母體平均值 μ 之位置集中。

樣本和的平均值$\bar{s} = E(S) = E(x_1 + x_2 + x_3 + \ldots + x_n) = E(x_1) + E(x_2) + E(x_3) + \ldots + E(x_n) = \mu + \mu + \mu + \ldots + \mu = n \times \mu$

無限基本單位數量之母體、($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本和 $s = x_1 + x_2 + x_3 + \ldots + x_n = \sum_{i=1}^{n} x_i$。

樣本和 S 的變異數σ_s^2與標準(偏)差 σ_s

$$\sigma_s^2 = Var(s) = V(S) = V(x_1 + x_2 + x_3 + \ldots + x_n) = V(x_1) + V(x_2) + V(x_3) + \ldots + V(x_n) = \sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2 = n \times \sigma^2$$

$$\sigma_s = \sqrt{n \times \sigma^2} = \sigma \times \sqrt{n}$$

有限基本單位數量之母體(抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本和 S 的變異數σ_s^2與標準(偏)差 σ_s

$$\sigma_s^2 = n \times \sigma^2 \times \frac{N-n}{N-1}$$

$$\sigma_s = \sqrt{n \times \sigma^2 \times \frac{N-n}{N-1}} = \sigma \times \sqrt{n \times \frac{N-n}{N-1}}$$

樣本比率$\bar{p}$分布的期望值 $E(\bar{p})$或平均值$\mu_{\bar{p}}$即等於原始資料間斷隨機變數 X 母體比例 p

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

無限基本單位數量之母體、($\frac{n}{N} \le 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p \times q}{n} = \frac{p \times (1-p)}{n}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p \times q}{n}} = \sqrt{\frac{p \times (1-p)}{n}}$$

有限基本單位數量之母體(抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n}\times\frac{N-n}{N-1} = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p\times q}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times q}{n}}\times\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}}\times\sqrt{\frac{N-n}{N-1}}$$

樣本數量與成功機率較高者($n\times p > 5$ 和 $n\times q > 5$)

樣本比率$\bar{p}$分布的期望值 $E(\bar{p})$、平均值$\mu_{\bar{p}}$即等於母體比例 p。$E(\bar{p}) = \mu_{\bar{p}} = p$

無限基本單位數量之母體、($\frac{n}{N}\leq 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n} = \frac{p\times(1-p)}{n}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p\times q}{n}} = \sqrt{\frac{p\times(1-p)}{n}}$$

樣本比率$\bar{p}$的抽樣分布接近於**常態分布**，即$\bar{p}\sim N(p, \frac{p\times q}{n})$

有限基本單位數量之母體或($\frac{n}{N} > 0.05$) (抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n}\times\frac{N-n}{N-1} = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p\times q}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times q}{n}}\times\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}}\times\sqrt{\frac{N-n}{N-1}}$$

樣本數量與成功機率較低者($n\times p < 5$ 或 $n\times q < 5$)

樣本比率$\bar{p}$分布的期望值 $E(\bar{p})$、平均值$\mu_{\bar{p}}$即等於母體比例 p

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

無限基本單位數量之母體、($\frac{n}{N}\leq 0.05$)或有限基本單位數量之母體(抽出的樣本放回，以供下一個樣本抽取)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n} = \frac{p\times(1-p)}{n}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p\times q}{n}} = \sqrt{\frac{p\times(1-p)}{n}}$$

樣本比率$\bar{p}$的抽樣分布屬於**二項分布**，即$\bar{p}\sim$二項分布$(p, \frac{p\times q}{n})$

有限基本單位數量之母體或($\frac{n}{N} > 0.05$) (抽出的樣本不放回，以供下一個樣本抽取，故 x_1 和 x_2 抽取機率不同)

樣本比率$\bar{p}$的變異數$\sigma_{\bar{p}}^2$與標準(偏)差$\sigma_{\bar{p}}$

$$\sigma_{\bar{p}}^2 = Var(\bar{p}) = V(\bar{p}) = \frac{p\times q}{n}\times\frac{N-n}{N-1} = \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p\times q}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times q}{n}}\times\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}\times\frac{N-n}{N-1}} = \sqrt{\frac{p\times(1-p)}{n}}\times\sqrt{\frac{N-n}{N-1}}$$

樣本比率$\bar{p}$的抽樣分布屬於超幾何分布，即$\bar{p}\sim$超幾何分布$(p, \frac{p\times(1-p)}{n}\times\frac{N-n}{N-1})$