上奇資訊股份有限公司
Grandtech Information Co., Ltd.

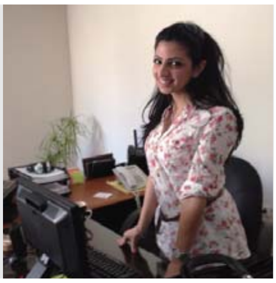
Silicon Stone

www.onlinetest.com.tw

全球著名學者與實務專家組成Data Analyst的命題委員，經過不斷的開會，有妥協，有衝突
最後將此系列證照分三個方向，Statistical Knowledge Today、Data Power Today 和Data Analysis Using Excel
相信經由這個證照訓練，未來各位可在這大數據時代，享受數據資訊。

計分方式 Exam Scoring

本考科共40題，選擇題有35題，複選題有5題，每題有25分，滿分是1000分，700分為及格分數。



測驗內容及科目

Statistical Knowledge Today 國際證照

- 統計學觀念及術語
- 資料描述
- 基礎圖表
- 機率
- 常見分配模型
- 抽樣分配與信賴區間估計

Data Power Today 國際證照

- 假設檢定
- 變異數分析
- 類別資料卡方檢定
- 簡單線性迴歸分析
- 複線性回歸分析
- 時間數列分析

Data Analysis Using Excel 國際證照

- 原始/群組資料與統計圖表
- 各種分配機率與切點求取
- 抽樣與亂數產生器
- 區間估計與假設檢定
- 類別資料卡方檢定
- 二維資料相關分析
- 迴歸分析與模型建構

證照樣式 License Sample

QR Code
辯識認證資格

通過測驗後，本公司將寄出姓名為中英文之合格證書，請至考試中心領取。

台灣獨家授權代理經銷商

上奇資訊 服務專線：02-2562-7969 傳真專線：02-2562-5269
http://www.grandtech.info/

佳魁資訊 服務專線：02-2562-7756 傳真專線: 02-2562-7716
http://www.topteam.cc /

Join Now

104 台北市中山區南京東路二段98號8樓之1

Certificate of Completion

To the Friends of learning everywhere, Greeting:

Be it known that

Data Analyst

Joshiua Karl Fish

having completed academic examination of the

Data Analyst

Sep., 15, 2014

Silicon Stone Education, Inc.

Silicon Stone

Serial Number :

Certificate Query Website : http://www.siliconstone.com
To complete the certificate query process, please visit this website and enter your serial number.

目 錄

CONTENTS

1 導 論

2 敘述統計

3 機率論基礎

4 機率分配模型

5 抽樣分配

6 信賴區間估計

單　元

1

導論

本認證教科書主要是一引領大家在認識或熟悉統計實務應用之前的相關基礎建立教材。隨著科技的進步，資料的取得已相對容易許多，不過資料如果沒有經過整理與萃取，它就只是一堆數字，對決策沒有太大的助益。但是如果資料能做有效的處理，它就會被轉成有用的資訊，進而協助決策者來提升決策的品質。

在本單元中，我們首先將探討何謂統計學？接著介紹幾個重要及常用的統計名詞、資料分類及抽樣方法，以利後續課程內容的了解。

在介紹統計相關名詞之前，首先讓我們先定義何謂統計學？

統計學 (Statistics) 為資料處理與分析的科學。包含：收集資料、資料如何的分類、處理及組織資料、總結資料、分析資料、解析資料並作適當的推論。

I. 統計重要及常用名詞

1. 母體 (Population)

所謂母體是指研究者關心或感興趣的一些項目所成之集合。一般而言母體的範圍是相當大甚至無限大，它的組成元素 (Element) 通常是指人或受測體。例如“某大學全體學生”或“台灣小學的學生”等。

2. 樣本 (Sample)

所謂樣本是指從母體中所抽取出來的一組元素。

3. 參數或母數 (Parameter)

用來描述母體特徵的值，稱之為參數或母數。例如“某大學全體學生的平均身高 (μ)”。

4. **統計量** (Statistic)

透過樣本資料所計算得到且用來描述樣本特徵的值，稱之為統計量。

5. **描述性統計** (Descriptive Statistics)

舉凡使用統計圖表或統計數字來描述資料概況的方法，皆被泛稱為描述性統計。

6. **推論性統計** (Statistical Inference)

舉凡透過樣本資料來對母體進行估計、預測及下決策的過程，皆被泛稱為推論性統計。

II. 資料型態

一般而言資料會被分類成數量的 (Quantitative 或 Numerical) 或質性的 (Qualitative 或 Categorical)，如表 1.1 所示。

表 1.1 資料的型態

資料	
質性資料 Qualitative 或 Categorical	數量資料 Quantitative 或 Numerical
例如：性別 宗教信仰 星座 血型	例如：體重 家庭年收入 速度 距離

有時候資料也會被分成名目尺度 (Nominal Scale)、順序尺度 (Ordinal Scale) 區間尺度 (Interval Scale) 及比率尺度 (Ratio Scale) 等四種不同尺度來探討；名目尺度的例子有婚姻狀況、宗教信仰、血型及星座等，另外像是電影分級雖然是分類資料的一種，但資料依其情色或暴力等情節的程度不同被分成普遍級、輔導級、限制級的順序，所以是屬於順序尺度的例子。至於區間尺度和比率尺度的差異只在於是否存在真實原點，區間尺度的例子有溫度等，比率尺度的例子有體重、家庭年收入、速度、距離等。

III. 常見的機率抽樣方法

抽樣方法依其抽樣方式可大致分成下列四種方法：

1. 簡單隨機抽樣 (Simple Random Sample)
2. 分層抽樣 (Stratified Sample)
3. 群集抽樣 (Clustered Sample)
4. 系統抽樣 (Systematic Sample)

演練題 (Q & A)

一、是非題

1 ：根據樣本統計量 (Statistic) 對於母體母數 (Parameter) 作推論或者估計其信賴區間是為統計推論。

Ans O

2 ：描述樣本變數的總結數稱為統計量 (Statistic)。

Ans O

3 ：描述母體 (Population) 變數的總結數我們稱之為母數 (Parameter)。

Ans O

4 ：名目尺度 (Nominal Scale) 資料如果具有順序性質就可成為了順序尺度 (Ordinal scale) 資料。

Ans O

5 ：區間尺度 (Interval Scale) 資料必定具備了有絕對 0 點可以做相對的比較。

Ans X

6 ：對於所研究個體間隨著量測會產生變化的特徵稱之為變數。

Ans O

17 ：統計學所使用的資料通常區分為量化資料與質化資料，並可以對應使用 4 種尺度加以處理。

Ans O

8 ：任何資料都可歸類為比率尺度 (Ratio Scale) 來進行資料處理。

Ans X

9 ：我們經由收集、整理資料、進行分類、繪製圖表並總結資料這些都是敘述統計的範疇。

Ans O

二、選擇題

1 ：根據樣本統計量 (Statistic) 對於母體母數 (Parameter) 作推論或者估計其信賴區間是為

A：樣本 (Sample)

B：統計學 (Statistics)

C：統計推論

D：敘述統計

Ans C

2 ：年薪的最佳量測尺度為

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans D

3 ：我們想知道大學畢業同學的年薪資為何？因此隨機抽取了 100 位大學畢業同學做研究。請問此研究中之母體 (Population) 為何？

A：所有的大學畢業同學

B：抽取的 100 位大學畢業同學

C：大學生的年薪資

D：100 位大學生的平均年薪

Ans A

4 ：我們想知道大學畢業同學的年薪資為何？因此隨機抽取了 100 位大學畢業同學做研究。請問此研究中之樣本為何？

A：所有的大學畢業同學

B：抽取的 100 位大學畢業同學

C：大學生的年薪資

D：100 位大學生的平均年薪

Ans B

5 ：我們想知道大學畢業同學的年薪資為何？因此隨機抽取了 100 位大學畢業同學做研究。請問此研究中之變數為何？

A：所有的大學畢業同學

B：抽取的 100 位大學畢業同學

C：大學生的年薪資

D：100 位大學生的平均年薪

Ans C

6 ：我們想知道大學畢業同學的年薪資為何？因此隨機抽取了 100 位大學畢業同學做研究。請問此研究中的統計量 (Statistic) 為何？

A：所有的大學畢業同學

B：抽取的 100 位大學畢業同學

C：大學生的年薪資

D：100 位大學生的平均年薪

Ans D

7 ：我們想知道大學畢業同學的年薪資為何？因此隨機抽取了 100 位大學畢業同學做研究。請問此研究中之母數 (Parameter) 為何？

A：所有的大學畢業同學的平均年薪

B：抽取的 100 位大學畢業同學

C：大學生的年薪資

D：100 位大學生的平均年薪

Ans A

8 ：以下何者為量化資料？

A：性別

B：血型

C：郵遞區號

D：身高

Ans D

9 ：以下何者為質化資料？

A：身高

B：體重

C：血型

D：年齡

Ans C

10 ：任何資料均能夠被視為何種尺度以進行統計分析？

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans A

11 ：郵遞區號在資料處理上是屬於何種尺度？

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans A

12 ：何種尺度資料具備有絕對 0 點可供相對比較分析？

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans D

13 ：某研究問題紀錄了目前的室內溫度、例如 :20 度 C、40 度 C。試問處理此問題所使用的最佳尺度為何？

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans C

14 ：某問卷問題的選答項計有 : 非常喜歡、喜歡、普通、不喜歡、非常不喜歡。試問處理此問題所使用的最佳尺度為何？

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

Ans B

三、複選題

1 ：請問下列何者屬於質化 (類別) 變數：

A：血型

B：性別

C：身高

D：體重

E：星座

Ans A, B, E

2 ：請問下列何者屬於量化 (數值) 變數：

A：血型

B：年齡

C：性別

D：體重

E：年薪

Ans B, D, E

3 ：我們處理年齡資料可以使用的尺度有

A：名目尺度 (Nominal Scale)

B：順序尺度 (Ordinal Scale)

C：區間尺度 (Interval Scale)

D：比率尺度 (Ratio Scale)

E：常數尺度 (Constant Scale)

Ans A, B, C, D

4 ：以下何種軟體經常使用來進行敘述統計分析？

A：SAS

B：SPSS

C：EXCEL

D：WORD

E：PPT

Ans A, B, C

5 ：下列何者為使用抽樣調查的原因？

A：樣本可能遭破壞

B：可以得到完全正確的結果

C：節省時間

D：節約經費

E：無誤差的產生

Ans A, C, D

單　元

2

敘述統計

如同在第一單元所介紹，母體的範圍是相當大甚至是無限大的，在有限的資源底下，想要完全了解母體的特徵，幾乎是不太可能完成的艱辛任務。統計的使命即是透過有限的 (樣本) 資料儘可能”正確”的揭發出母體的全貌，所以資料就成了統計中最重要的元素，而如何讓資料顯現其內在涵義，則為高深且值得追求的學問。

I. 資料整理及統計圖表

最快及有效整理資料的方式即統計圖表，次數分配表是最基本的，統計圖則是另一種較生動的呈現方式，常用的統計圖有直條圖或長條圖 (Bar Chart)、圓餅圖 (Pie Chart)、直方圖 (Histogram)、莖葉圖 (Steam-and-Leaf Display)、箱形圖 (Box Plot) 以及時間序列圖 (Time Plot)。接下來，我們將介紹幾種針對不同型態的資料，可以採用統計圖表來進行初步的資料整理與彙總。一般而言，區間尺度資料多建議採用次數分配表 (Frequency Distribution Table) 、直方圖及莖葉圖等來呈現資料分佈狀況。不過有特殊目的時，亦可以採用箱型圖來呈現資料分佈情形。至於順序尺度資料，一般多將之視為分類資料，可採用直條圖 (長條圖) 或圓餅圖來描述資料。質性資料則建議採用直條圖或圓餅圖來整理資料，另外時間數列資料多以時間序列圖來描述資料隨時間變動的走勢。

最快速且有效整理資料的方式之一是次數分配表。編製次數分配表的步驟如下：

Step 1：決定組數 k

Step 2：決定組距 (Class Width)，找出資料的最大值 $X_{(Max)}$ 和最小值 $X_{(Min)}$

$$\text{組距} = \frac{X_{(Max)} - X_{(Min)}}{k}$$

Step 3：設定組界 (Class Limits)，通常是從最小值開始依序加上組距即可求得，有時候為了避免資料落在組界上，造成分類的尷尬，可以比最小值稍小的數來當第一組的下界 (例如：$X_{(Min)} - 0.5$)

Step 4：記錄各組出現的頻率 (Frequency)

如此便可完成次數分配表。

隨著電腦技術的進步，統計圖的繪製雖然說相對簡單，但統計圖的主要目的其實是要整理資料，並希望從中獲取精簡且有用的資訊，但是一般人卻不一定可以使用合宜的統計圖來呈現及萃取資訊。下表 2.1 是我們依照資料的型態所建議採用的圖形方法。

表 2.1　圖形表示法的選擇

資料	
質性資料 Qualitative (或 Categorical)	數量資料 Quantitative (或 Numerical)
例如：直條圖、圓餅圖等。	例如：直方圖、莖葉圖、圓形圖、肩形圖 (Ogive)、時間序列圖、箱型圖等。

II. 敘述統計：集中趨勢及離散情形

本小節主要是延續描述性統計的討論，介紹處理複雜資料的一些技巧以及如何透過簡單的總結數字來表達資料中所隱含的訊息。一般來說，資料有可能是完全未經處理的原始資料 (Raw Data) 或者是已經整理過的次級資料 (Secondary Data)。本小節將針對原始資料的處理方法加以介紹，而下一小節則將針對已分組好的次級資料或稱為分組資料 (Group Data) 來進行探討。首先我們將介紹幾個重要的名詞，對數量資料集我們多會研究二個基本性質，一個是集中趨勢 (Central Tendency)，另一個是離散情形 (Variability 或 Dispersion)。

通常資料會群聚在中央值附近，這樣的模式就被稱為集中趨勢，而常被用來測量集中趨勢的方法有下列三種：1. 眾數 (Mode) 2. 中位數 (Median) 3. 平均數 (Mean)，他們的定義分別如下：

1. 眾數：在給定資料集中，發生次數最多的值就叫做眾數。

範例 2.1 考慮下列資料集，求該資料集的眾數。

-2，0，5，5，4，3，7

Ans 因為 5 出現的次數 (2 次) 最多，所以在此資料集當中的眾數為 5。

2. 中位數：簡單地說在一組排序的資料中，將資料切成下 50% 及上 50% 的值 (或是最中間的值) 即為中位數，更明確的定義如下：

先將資料集中之 n 個觀察值進行排序，簡記作 $X_{(1)}$，$X_{(2)}$，…，$X_{(n)}$，其中 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，則對應最中間位置的數即為中位數，所以

(i) 若觀察值的總個數是奇數，則中位數 = $X_{((n+1)/2)}$

(ii) 若觀察值的總個數是偶數，則中位數 = $\dfrac{X_{(n/2)} + X_{(n/2+1)}}{2}$

範例 2.2 考慮下列資料集 (n = 7)，求該資料集的中位數。

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$
200	200	250	275	295	300	350

Ans 中位數 = $X_{((7+1)/2)} = X_{(4)} = 275$

3. 平均數

(i) 母體平均數，μ

給定母體資料集，所有觀察值 X_1，X_2，⋯，X_N 的總和除以母體總個數 (N)，即稱為母體平均數。平均數的定義如下：

$$\mu = \sum_{i=1}^{N} X_i / N$$

(ii) 樣本平均數，$\overline{X}$

在第一單元中，我們介紹了參數及統計量，其中母體的特徵 (例如：母體平均數、母體變異數) 稱為參數。而樣本的特徵 (例如：樣本平均數、樣本變異數) 稱之為統計量。因為母體資料範圍太大，所以不易取得參數值，一般的做法是透過樣本來推估母體的特徵。所以樣本平均數 $\overline{X}$ 就是用來估計母體平均數 μ 的統計量，定義如下：

$$\overline{X} = \sum_{i=1}^{n} X_i / n \text{，其中 } n \text{ 是樣本數}$$

範例 2.3 考慮下列資料集，求該資料集的平均數。

10，20，30，40，50

Ans 平均數 $= \dfrac{10+20+30+40+50}{5} = 30$

測量集中趨勢的測量值尚有全距中心 (Midrange)，主要是來測量資料距離長度的中心，其計算公式為資料最大值加最小值後除以 2。雖然眾數、中位數與平均數都是衡量資料集的集中趨勢，但不一定三者衡量出來的結果皆相同。因此得視動機來選擇適合的測量方式。眾數是三者中最容易計算的，但眾數並不唯一，也因此造成不易偵測出資料的中心點，這是它方法上的缺點。另外，平均數較中位數易受極

端值的影響，所以資料集當中若存在著極端值，中位數是較合適用來衡量中心位置的一種方法。

資料分布形狀的不同，亦會對平均數及中位數大小關係產生影響。在圖 2.1 中，可以看出當資料集呈現右偏時，平均數會大於中位數，而資料集呈現左偏的時候，平均數會小於中位數。若資料呈現對稱，則平均數會等於中位數。

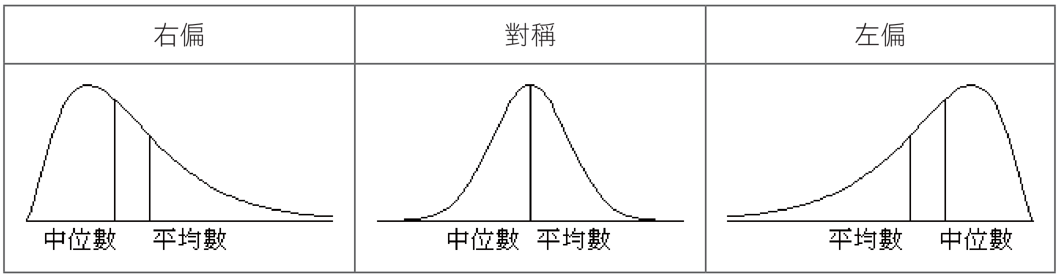


圖 2.1 不同分佈下，平均數與中位數之比較。

在集中趨勢中，我們討論了三種衡量方法，但縱使兩個資料集彼此之間的眾數、中位數及平均數皆相等，但其他方面的差異仍然可以是很大的。而用來衡量離散情形的量數有全距 (Range)、變異數 (Variance)、標準差 (Standard Deviation)、變異係數 (Coefficient of Variation)、四分位數 (Quartile) 、百分位數 (Percentile)、四分位距 (Interquartile Range) 等。下面就逐一針對較重要的量數來介紹他們的定義：

1. 全距

資料集中最大觀察值 ($X_{(n)}$) 減掉最小觀察值 ($X_{(1)}$) 即為全距，也就是

$$全距 = X_{(n)} - X_{(1)}$$

範例 2.4 考慮下列資料集，求該資料集的全距。

2，4，6，8，10，11，12，14

Ans 全距 = 14 – 2 = 12

2.　變異數

(i) 母體變異數，σ^2

對整個母體的資料集而言，變異數主要是衡量所有觀察值與中心點 (母體平均數 μ) 差異的平均平方和。定義如下：

$$\sigma^2 = \sum_{i=1}^{N}(X_i - \mu)^2 / N$$

其中 N 是母體內元素個數，μ 是母體平均數。

(ii) 樣本變異數，S^2

樣本變異數 S^2 就是用來估計母體變異數 σ^2 的統計量。定義如下：

$$S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 /(n-1)$$

其中 n 是樣本數，$\overline{X}$ 是樣本平均數

範例 2.5 考慮下列資料集 (n = 5)，求該資料集的樣本變異數。

0，5，2，4，4

Ans 我們將資料整理如下表 (1)~(5) 欄，首先將資料 X_i 填入

(1)	(2)	(3)	(4)	(5)
i	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	X_i^2
1	0	-3	9	0
2	5	2	4	25
3	2	-1	1	4
4	4	1	1	16
5	4	1	1	16
$\sum$	15		16	61

再利用第 (2) 欄的加總求出

$$\overline{X}=\frac{\sum_{i=1}^{n}X_{\mathrm{i}}}{n}=\frac{15}{5}=3$$

之後再將各觀察值 X_{i} 減去 $\overline{X}$ 得第 (3) 欄，再分別平方後得第 (4) 欄加總，再帶入 S^2 的公式；$S^2=\dfrac{\sum_{i=1}^{n}(X_{\mathrm{i}}-\overline{X})^2}{n-1}=\dfrac{16}{5-1}=4$，即可求得樣本變異數等於 4。

另外一個作法是利用較簡便的計算公式，所以須準備好表格中的欄 (1)、(2) 及 (5)，再帶入下列公式

$$S^2=\frac{\sum_{i=1}^{n}X_{\mathrm{i}}^2-n\overline{X}^2}{n-1}=\frac{61-5\times(3)^2}{5-1}=4$$

而得到樣本變異數的值 4。

3. 標準差

變異數雖然可以衡量出資料距離中心點的離散情形，但因為其單位和原始觀察值的單位不同 (是原始觀察值單位的平方)，使得解釋資料的困難度提高。所以一般人多以標準差來衡量離散度。它的定義如下：

(i) 母體標準差 $\sigma=\sqrt{\sigma^2}$

(ii) 樣本標準差 $S=\sqrt{S^2}$

範例 2.6 考慮下列資料集 (n = 5)，求該資料集的樣本標準差。

0，5，2，4，4

Ans 樣本標準差 = 2

4. 四分位數

將資料集分成 4 等份的三個數值即為四分位數，其中第 2 個四分位數 Q_2 即為中位數 (或者說是第 50 個百分位數 (Percentile))，而第一個四分位數 Q_1 (或第 25 個百分位數) 即為前半段 (前 50%) 資料的中位數，而第三個四分位數 Q_3 (或第 75 個百分位數)，即為後半段 (後 50%) 資料的中位數。

註

將資料集等分為 100 份的數值，即所謂的百分位數 (Percentile)。估計百分位數的方法有很多種，在此僅提供一種較常用的方式供大家參考。

假設一組含 $X_1, X_2, \cdots, X_n$ 觀察值的資料集，將觀察值排序後得 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，其第 p 個百分位數的定義為

$$P_p = \begin{cases} X_{([pn/100]+1)} & ,\text{若} pn/100 \text{不為整數} \\ \dfrac{X_{(pn/100)} + X_{(pn/100+1)}}{2} & ,\text{若} pn/100 \text{為整數} \end{cases}$$

註

Excel 所採用百分位數的估計方式和大部份統計學課本的定義並不相同，它的作法如下：首先，令 $1 + p(n - 1)/100 = k + d$，其中 k 為整數部份，d 為小數部份，則第 p 個百分位數為 $P_p = X_{(k)} + d(X_{(k+1)} - X_{(k)})$。

5. 四分位距，*IQR*

涵蓋中間 50% 資料的範圍稱為四分位距，它的定義如下：

$$IQR = Q_3 - Q_1$$

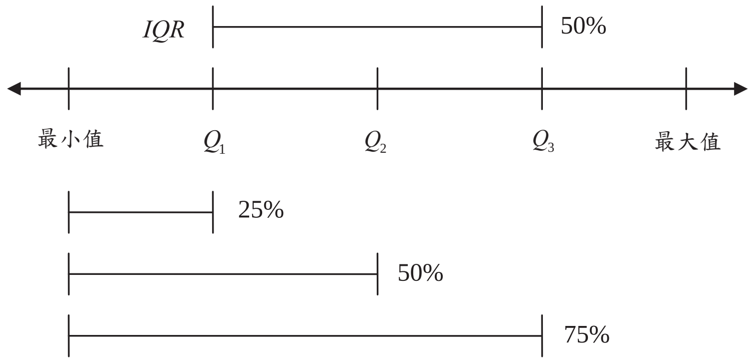


圖 2-2：四分位數 Q_1，Q_2，Q_3 和四分位距 IQR 的示意圖

範例 2.7 考慮下列資料集 (n = 10) ，求該資料集的全距、四分位數和四分位距。

-1，6，8，0，1，3，9，10，5，4

Ans 先將資料重新排序得

-1，0，1，3，4，5，6，8，9，10

所以　全距 = 最大值－最小值 = 10 － (-1) = 11

另外，$Q_2 = \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2} = \dfrac{4+5}{2} = 4.5$

前半段資料為 -1，0，1，3，4，所以 $Q_1 = 1$

後半段資料為 5，6，8，9，10 ，所以 $Q_3 = 8$

所以四分位距 $IQR = Q_3 - Q_1 = 8 - 1 = 7$

註

除了前述用中位數的方式求 Q_1、Q_3 之外，也可以用 P_{25} 和 P_{75} 來求 Q_1 及 Q_3。作法如下：

$$P_{25} = X_{([2.5]+1)} = X_{(3)} = 1$$
$$P_{75} = X_{([7.5]+1)} = X_{(8)} = 8$$

III. 群組資料的整理與歸納

本節主要是介紹如何整理及呈現群組資料，不論資料是初級或是次級資料，我們多是希望可以找到兩個描述資料集分佈狀況的總結數字或測量值。若資料只能透過次數的方式呈現，那麼這樣的資料集就被稱為群組資料集。為了方便下面公式的介紹，我們先將其形式及符號歸納成下列的表格：

序號	組別	次數	相對次數	累積相對次數	組中點
1	$L_0 \sim L_1$	f_1	r_1	c_1	m_1
2	$L_1 \sim L_2$	f_2	r_2	c_2	m_2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
k	$L_{k-1} \sim L_k$	f_k	r_k	c_k	m_k
總合		n	1		

對群組資料集而言，其主軸概念是使用組中點來代表對應組別中的觀察值。下面分別詳述之。

1. 集中趨勢

1-1. 眾數

發生次數最多的那個組別中所對應的組中點，即為群組資料集的眾數。

範例 2.8 某位醫生針對 20 名患有心臟病的病患進行檢查，得到其膽固醇指數的測量值如下，求群組資料集的眾數。

序號	膽固醇指數 (組別)	次數	組中點
1	190~200	2	195
2	200~210	4	205
3	210~220	5	215
4	220~230	7	225
5	230~240	2	235

Ans 在 5 個組別中，第四組發生的次數 (7 次) 最多

∴ 第 4 組的組中點 225 為眾數。

1-2. 中位數

因為群組資料集的中位數算法和後面四分位數中的第 2 四分位數 Q_2 相同，所以留待後面章節介紹。

1-3. 平均數

群組資料集平均數的定義如下：

$$\overline{X} = \frac{m_1 \times f_1 + m_2 \times f_2 + \cdots + m_k \times f_k}{n}$$

其中 f_i 為觀察值落在第 i 組的次數，m_i 為第 i 組的組中點 $i = 1,2,...,k$ ，n 為總觀察次數。

範例 2.9 以範例 2-8 的資料集，求群組資料集的平均數

Ans $\overline{X} = \dfrac{2\times195+4\times205+5\times215+7\times225+2\times235}{20} = 216.5$

2. 離散情形

2-1. 全距

在群組資料中，全距的定義是第一組下界 (L_0) 到最後一組上界 (L_k) 的距離。

即

$$全距 = L_k - L_0$$

範例 2.10 以範例 2-8 的資料集，求群組資料集的全距。

Ans 全距 = 240 － 190 = 50

2-2. 變異數

群組資料集變異數的定義為

$$S^2 = \frac{\sum_{i=1}^{k}\left[(m_i - \overline{X})^2 \times f_i\right]}{n-1}$$

其中 f_1 為觀察值落在第 i 組的次數，m_i 為第 i 的組中點 $i = 1,2,...,k$ ，k 為總觀察次數，$\overline{X}$ 為群組資料的平均數。

2-3. 標準差

群組資料集的標準差為變異數的正平方根，即 $S = \sqrt{S^2}$ 。

範例 2.11 以範例 2-8 的資料集，求變異數及標準差

Ans

序號	組別	f_1	m_i	$m_i - \bar{x}$	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 \times f_i$
1	190-200	2	195	-21.5	462.25	924.5
2	200-210	4	205	-11.5	132.25	529
3	210-220	5	215	-1.5	2.25	11.25
4	220-230	7	225	8.5	72.25	505.75
5	230-240	2	235	18.5	342.25	684.5
總合						2655

$$s^2=\frac{2655}{20-1}\cong 139.737$$

$$s=\sqrt{s^2}\cong 11.821$$

2-4. 百分位數

群組資料集的第 k 個百分位數 P_k 的定義如下：

$$P_k=L_{j-1}+\frac{(k/100-C_j)}{r_j}\times l$$

其中 L_{j-1} 為包含 P_k 這組 (j) 的下限，r_j 為這組的相對次數，C_j 為以 L_{j-1} 為上限的組 ($j-1$) 所累積之相對次數，l 為組距。

2-5. 四分位數及四分位距

群組資料跟第二小節原始資料一樣，將資料集分成 4 等份的三個數值即為四分位數，只是其中第 1 個四分位數 Q_1(或第 25 個百分位數)、第 2 個四分位數 Q_2 即為中位數 (或者說是第 50 個百分位數)、第 3 個四分位數 Q_3(或第 75 個百分位數) 的計算方式，是參照百分位數 P_k 的公式分別代入 k = 25，50，75 即可求出四分位數。至於涵蓋中間 50% 資料的範圍亦稱為內四分位距，它的定義如下：

$$IQR=Q_3-Q_1$$

範例 2.12 以範例 2-8 的資料集，求該群組資料集的中位數 (第 2 個四分位數)、第 1 個四分位數及第 3 個四分位數。

Ans

序號	組別	f_1	r_i	C_i
1	190-200	2	0.1	0.1
2	200-210	4	0.2	0.3
3	210-220	5	0.25	0.55
4	220-230	7	0.35	0.9
5	230-240	2	0.1	1

中位數為第 50 個百分位數 P_{50}，$k = 50$，包含 P_{50} 這組的下限 $L_{j-1} = 210$，包含 P_{50} 這組的相對次數 $r_j = 0.25$，以 L_{j-1} 為上限的組所累積之相對次數 $C_j = 0.3$，組距 $l = 10$。

$$P_{50} = 210 + \frac{(0.5 - 0.3)}{0.25} \times 10 = 218$$

第 1 個四分位數為第 25 個百分位數 P_{25}，$k = 25$，包含 P_{25} 這組的下限 $L_{j-1} = 200$，包含 P_{25} 這組的相對次數 $r_j = 0.2$，以 L_{j-1} 為上限的組所累積之相對次數 $C_j = 0.1$，組距 $l = 10$。

$$P_{25} = 200 + \frac{(0.25 - 0.1)}{0.2} \times 10 = 207.5$$

第 3 個四分位數為第 75 個百分位數 P_{75}，$k = 75$，包含 P_{75} 這組的下限 $L_{j-1} = 220$，包含 P_{75} 這組的相對次數 $r_j = 0.35$，以 L_{j-1} 為上限的組所累積之相對次數 $C_j = 0.5$，組距 $l = 10$。

$$P_{75} = 220 + \frac{(0.75 - 0.55)}{0.35} \times 10 = 225.71$$

IV. 柴比雪夫 (Chebyshev) 定理及經驗法則

柴比雪夫定理：假設隨機變數 X 之有限期望值為 μ，標準差為 σ，$k > 0$，則

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$$

舉例來說，當 $k = 3$ 時，則至少有 8/9 的資料會發生在 $(\mu - 3\sigma, \mu + 3\sigma)$ 區間內。

經驗法則：假如資料之機率分配是近似於鐘形 (常態) 分配，則根據經驗法則

(a) $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68\%$

(b) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$

(c) $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.7\%$

也就是說，當資料分配是近似鐘形對稱分配的話，則大約有 99.7% 的資料會發生在 $(\mu-3\sigma, \mu+3\sigma)$ 區間內。

柴比雪夫定理的優點是其能應用於任何資料結構，而經驗法則只能應用在對稱的資料結構。

V. 兩變數關係之量測

一般而言，較常被採用來衡量兩變數之關係的工具有 (一) 散佈圖 (Scatter Plot)，(二) 相關分析 (Correlation Analysis)。下面就依序加以介紹之。

V-1. 散佈圖

散佈圖主要是透過圖形來檢查兩變數之間的關係。作法是在座標平面上描點，例如圖 2.3 中所示。

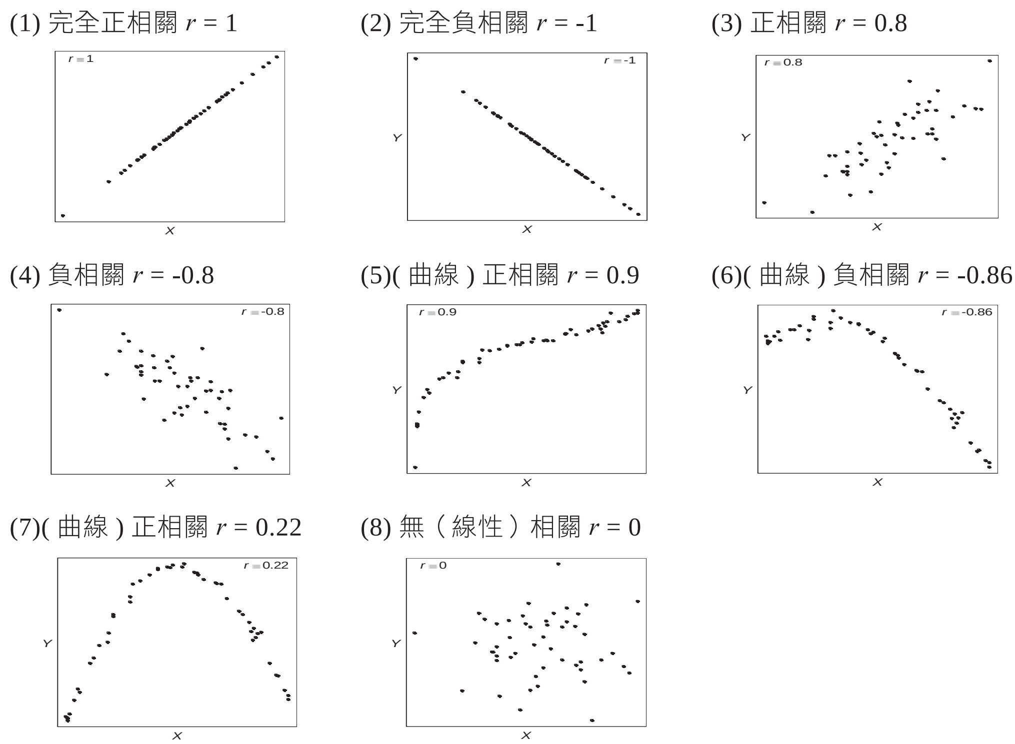


圖 2.3 各種散佈圖

從圖 2.3 (1)-(4) 可以看出 Y 和 X 之間約略呈線性關係，而圖 2.3 (5)-(7) 則顯示 Y 和 X 應存在著曲線的關係，圖 2.3 (7) 尤其明顯。但是透過圖 2.3 (8) 則很難判斷出 X 和 Y 之間函數關係的形式。雖然從散佈圖可以很快速得到兩變數間的”對應關係”，然而透過目測來下結論較易引起爭議，況且這些充其量只提供了方向，卻無法正確及有效的描述出兩變數間的函數關係與強度。

V-2. 相關係數，Correlation Coefficient

相關係數是另一個可以用來衡量兩個變數間相關性的指標，本節依母體及樣本分別介紹其對應之定義。

V-2-1. 母體相關係數 ρ，Population Correlation Coefficient

$$\rho = \frac{Cov(X,Y)}{\sqrt{VarX}\sqrt{VarY}}$$

其中 $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$ 為 X 和 Y 的共變異數 (Covariance)

$Var(X) = E(X - E(X))^2$ 為 X 的變異數

$Var(Y) = E(Y - E(Y))^2$ 為 Y 的變異數

一般而言，母體相關的參數都是未知的，因此須要藉由樣本來進行推估。相關係數並不例外，用來估計母體相關係數 ρ 的統計量為樣本相關係數 r，其定義如下。

V-2-2. 樣本相關係數 r ，Sample Correlation Coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$$

其中 $S_{XY} = \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$ 為 X 和 Y 的共變異 (Covariation)

1335-CC1402-Ch1-Ch3.indd 17 2014/10/22 上午 11:54:50

$$S_{XX} = \sum_{i=1}^{n} (X_i - \overline{X})^2 \text{ 為 } X \text{ 的變異}$$

$$S_{YY} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \text{ 為 } Y \text{ 的變異}$$

有時候為了方便計算，公式亦可被簡化如下：

$$S_{XY} = \sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i) / n$$

$$S_{XX} = \sum_{i=1}^{n} X_i^{\ 2} - (\sum_{i=1}^{n} X_i)^2 / n$$

$$S_{YY} = \sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2 / n$$

下面我們就用一個例子來說明如何求算相關係數。

範例 2.13 某經紀公司想了解優酪乳的銷售量是否和廣告宣傳費用有關，所以調查了 10 個檔次優酪乳的銷售量 (Y，百箱) 及這 10 個檔次優酪乳的宣傳費用 (X，單位：萬元)，得下列數據：

$$\overline{X} = 2.7 \qquad \overline{Y} = 7.6 \qquad S_{XX} = 230.5 \qquad S_{YY} = 20.9 \qquad S_{XY} = 67.8$$

1. 請計算樣本相關係數。請問兩者是否呈現正 (或負) 相關？
2. 請問根據調查的結果是否支持優酪乳檔次的宣傳費用愈高，優酪乳的銷售量亦較高的論點？

Ans

1. $$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}} \sqrt{S_{YY}}} = \frac{67.8}{\sqrt{230.5} \sqrt{20.9}} \approx 0.9768$$

   所以兩者呈現正相關。

2. 是的，調查的結果支持優酪乳檔次的宣傳費用愈高優酪乳的銷售量亦較高的論點。

在範例 2-13 中我們得到 $r = 0.9768$，這個數字代表什麼意義呢？一般而言，若 Y 值隨著 X 值變大而增大，則稱之為正相關 (例如：圖 2.3-(1)、(3)、(5)、(7))。反之，若 Y 值隨著 X 值變大而減少，則稱之為負相關 (例如：圖 2.3-(2)、(4)、(6))。另外，不是正相關亦不是負相關的情形就稱為無相關 (例如：圖 2.3- (8))。所以範例 2-13 中 r =0.9768 代表優酪乳檔次的宣傳費用和優酪乳的銷售量呈正相關。圖 2.3- (7) 雖然名義上為正相關，但線性強度微弱 ($|r| = 0.2$)。

演練題 (Q & A)

一、是非題

1 ：圓餅圖 (Pie Chart) 與直條圖 (Bar Chart) 通常是用來處理類別資料。

Ans O

2 ：肩形圖 (Ogive) 是將累積次數表或累積相對次數表的數值分組資料以圖形的方式表達。

Ans O

3 ：給予以下 11 筆樣本資料 : 3 4 4 5 5 5 6 6 6 6 6；則 IQR 四分位距 (Interquartile Range, IQR) 為 2。

Ans O

4 ：直方圖 (Histogram) 可以直接由圖上知道有多少次數或多少比例的觀察值低於某一特定值。

Ans X

5 ：總結表 (Summary-Table) 適用於質化資料，並可於表中列出所有分類及發生次數。

Ans O

6 ：從一母體中抽取出 11 個樣本，得到下列排序資料如下：2, 2, 2, 2, 3, 3, 4, 6, 8, 10, 13。則資料 $x = 13$ 的 Z-score = 2.1(小數後第二位四捨五入)。

Ans O

7 ：莖葉圖 (Stem and Leaf Display) 可用來做資料排序，但是不適用於大量資料呈現。

Ans O

8 ：從箱型圖 (Box Plot) 中可得知資料的平均數與眾數。

Ans X

9 ：我們可以直接由 Pareto 圖上知道有多少次數或多少比例的資料落在那些主要組別。

Ans O

10 ：散佈圖 (Scatter Plot) 適用於呈現兩類別資料間的關聯情形。

Ans X

11 ：Midquartile 是用來表達數值資料的集中趨勢。

Ans O

12 ：全距中心 (Midrange) 是用來表達數值資料的散佈情形。

Ans X

13 ：四分位距 (Interquartile Range, *IQR*) 是用來表達數值中心 50% 資料的散佈情形。

Ans O

14 ：給予以下 10 筆樣本資料：3 4 4 5 5 5 6 6 6 6；此資料為右偏。

Ans X

15 ：某次考試成績的總結如下：

Variable	Score	Mode	65	Q_1	60
N	50	St-dev	15.8	Q_3	78
Mean	?	Max	98	Sum	3350
Median	66	Min	42	Range	?

則此次考試成績為右偏。

Ans O

16 ：從箱型圖 (Box Plot) 上我們可以得知全距為何。

Ans O

二、選擇題

1 ：從一母體中抽取出 11 個樣本得到下列排序資料如下：2, 2, 2, 2, 3, 3, 4, 6, 8, 10, 13。試問樣本資料 $x = 2$ 的 Z-score 為何？

A：0.9

B：-1.9

C：-1.5

D：-0.8

Ans D

2 ：比例尺度資料通常可以繪製成以下何種統計圖？

A：圓餅圖 (Pie Chart)

B：直條圖 (Bar Chart)

C：直方圖 (Histogram)

D：Pareto 圖

Ans C

3 ：從一母體中抽取出 11 個樣本得到下列排序資料：2, 2, 2, 2, 3, 3, 4, 6, 8, 10, 13。試問樣本標準差為何？

A：5.1

B：4.9

C：4.2

D：3.8

Ans D

4 ：從一母體中抽取出 11 個樣本得到下列排序資料：2, 2, 2, 2, 3, 3, 4, 6, 8, 10, 13。試問樣本變異數為何？

A：14.4

B：144

C：12

D：13.1

Ans A

5 ：將一組數值量化資料製作成次數分配表不需要哪一個過程。

A：求全距

B：定組數

C：定組界

D：計算平均數

Ans D

6 ：我們可以從箱型圖 (Box Plot) 中得到下列哪一個統計量？

A：平均數

B：變異數

C：四分位距 (Interquartile Range)

D：眾數

Ans C

7 ：下列何者不是莖葉圖 (Stem and Leaf Display) 的特性？

A：可以用來做資料排序

B：莖數為表示相對次數，葉數表示分類

C：不適用於大量或複雜的資料

D：可以保持原始資料

Ans B

8 ：下列何者不是箱型圖 (Box Plot) 所要表達之特定點？

A：平均數

B：中位數

C：最大數

D：最小數

Ans A

9 ：下列何種圖表可以直接知道多少比例的觀察值低於某特定值？

A：散佈圖 (Scatter Plot)

B：肩型圖 (Ogive)

C：點圖 (Dot Plot)

D：直條圖 (Bar Chart)

Ans B

10 ：下列何種中心集中趨勢的測量值也適合用於類別資料變數之描述？

A：全距中心 (Midrange)

B：平均數

C：中位數

D：眾數

Ans D

11 ：根據經驗法則，99.7% 的鐘形分配母體資料會落在哪個區間內。

A：$[\mu-\sigma, \mu+\sigma]$

B：$[\mu-2\sigma, \mu+2\sigma]$

C：$[\mu-3\sigma, \mu+3\sigma]$

D：$[\mu-4\sigma, \mu+4\sigma]$

Ans C

12 ：根據柴比雪夫定律，至少 88.98% 的母體資料會落在哪一個區間內。

A：$[\mu-\sigma, \mu+\sigma]$

B：$[\mu-2\sigma, \mu+2\sigma]$

C：$[\mu-3\sigma, \mu+3\sigma]$

D：$[\mu-4\sigma, \mu+4\sigma]$

Ans D

13 ：從一母體中抽取出 11 個樣本得到下列排序資料：2, 2, 2, 2, 3, 3, 4, 6, 8, 10, 13。試問樣本中位數為何？

A：2

B：3

C：4

D：3.5

Ans B

14 ：給予以下 10 筆樣本資料 : 3 4 4 5 5 5 6 6 6 6; 其平均數為：

A：5.2

B：4.5

C：5.0

D：4.8

Ans C

15 ：給予以下 10 筆樣本資料 : 3 4 4 5 5 5 6 6 6 6; 其眾數為：

A：4.5

B：5

C：6

D：5.5

Ans C

16 ：給予以下 10 筆樣本資料 : 3 4 4 5 5 5 6 6 6 6; 其全距為：

A：3

B：4

C：5

D：4.5

Ans A

17 ：給予以下 11 筆樣本資料 : 3 4 4 5 5 5 6 6 6 6 6; 其全距中心 (Midrange) 為：

A：4

B：4.5

C：5

D：4.9

Ans B

18 ：假如 x, y 數值資料均落在 $Y = -0.5X$ 直線上，則此兩變數之線性相關係數 ρ 為何？

A：$\rho = 0$

B：$\rho = 1$

C：$\rho = -1$

D：$\rho = -0.5$

Ans C

19 ：下列何者為計算樣本變異數之公式？

A：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$

B：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$

C：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

D：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1}$

Ans C

20 ：下列何者為計算母體變異數之公式？

A：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$

B：$\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{N-1}$

C：$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

D：$\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$

Ans D

21 ：下列何者為計算兩變數 X, Y 樣本相關係數 r 之公式：

A：$r = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}$

B：$r=\dfrac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}$

C：$r=\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_i^2}$

D：$r=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$

Ans A

22 ：根據經驗法則，單峰對稱分配資料落在$[\mu-2\sigma, \mu+2\sigma]$之區間約有多少百分比。

A：99.7%

B：68%

C：80%

D：95%

Ans D

23 ：根據經驗法則，單峰對稱分配資料落在$[\mu-3\sigma, \mu+3\sigma]$之區間約有多少百分比。

A：68%

B：89%

C：99.7%

D：95%

Ans C

24 ：根據柴比雪夫定理，任何分配的資料落在$[\mu-3\sigma, \mu+3\sigma]$之區間至少有多少百分比的機率。

A：88.8%

B：90%

C：95%

D：99.7%

Ans A

25 ：某次考試成績的總結如下表：

Variable	Score	Mode	65	Q_1	60
N	50	St-dev	15.8	Q_3	78
Mean	?	Max	98	Sum	3350
Median	66	Min	42	Range	?

請問此次考試的平均成績為何？

A：67

B：66

C：65

D：78

Ans A

26 ：某次考試成績的總結如下表：

Variable	Score	Mode	65	Q_1	60
N	50	St-dev	15.8	Q_3	78
Mean	?	Max	98	Sum	3350
Median	66	Min	42	Range	?

請問此次考試成績的全距為何？

A：50

B：42

C：56

D：60

Ans C

27 ：某次考試成績的總結如下表：

Variable	Score	Mode	65	Q_1	60
N	50	St-dev	15.8	Q_3	78
Mean	?	Max	98	Sum	3450
Median	66	Min	42	Range	?

請問此次考試成績的四分位距 (Interquartile Range) 為何？

A：69

B：22

C：18

D：15.8

Ans C

28 ：給予下列成績的分組總結資料表。

成績	次 數
0 ~ 25	10
25 ~ 50	20
50~ 75	30
75 ~ 100	40

請問此分組資料的平均數為何？

A：75

B：70

C：62.5

D：50

Ans C

29 ：給予下列成績的分組總結資料表。

成績	次 數
0 ~ 25	10
25 ~ 50	20
50~ 75	30
75 ~ 100	40

請問此分組資料的中位數約為何？

A：66.7

B：75

C：50

D：55

Ans A

30 ：給予下列成績的分組總結資料，請問大約及格的人數為何：

成績	人數
0 ~ 25	10
25 ~ 50	20
50~ 75	30
75 ~ 100	40

A：50

B：62.5

C：58

D：70

Ans C

三、複選題

1 ：類別資料通常可以繪製成以下何種統計圖？

A：直方圖 (Histogram)

B：莖葉圖 (Stem and Leaf)

C：時間序列 (Time Series)

D：圓餅圖 (Pie Chart)

E：直條圖 (Bar Chart)

Ans D, E

2 ：比率尺度 (Ratio Scale) 資料通常可以繪製成以下何種統計圖？

A：直方圖 (Histogram)

B：莖葉圖 (Stem and Leaf)

C：箱型圖 (Box Plot)

D：圓餅圖 (Pie Chart)

E：直條圖 (Bar Chart)

Ans A, B, C

3 ：箱型圖 (Box Plot) 上有表達下列的那些資料點？

A：第三四分位數

B：中位數

C：最大值

D：眾數

E：平均數

Ans A, B, C

4 ：以下哪些是描述數值資料的集中趨勢？

A：平均數

B：眾數

C：變異數

D：中位數

E：全距中心 (Midrange)

Ans A, B, D, E

5 ：從箱型圖 (Box Plot) 上我們可以得知那些資料分配的情形？

A：全距

B：四分位距 (Interquartile Range)

C：標準差

D：變異數

E：平均數

Ans A, B

單　元

3

機率論基礎

在日常生活上我們常常會碰到機率或是機會的名詞，例如，

1. 一個資本家聲稱：一個剛開始的網路公司在一年後只有 5% 存活的機會。
2. 一位槍砲專家說：某特定迫擊砲當它擊發時有 0.03 爆炸失敗的機率。

首先我們要澄清的是機率和機會基本上是相同的事，機率是以 0 到 1 之間的數字來表示，當我們用機率乘以 100%，所得的值我們稱為機會。因此，機會是以 0% 到 100% 間的數字來表示。

我們注意到在日常生活環境中每件事情都是不確定的，馬克吐溫 (Mark Twain) 的名言說，“除了死亡和稅，每件事情都是不確定的 (everything is uncertain except death and tax)”。不管一個網路公司在一年後存活與否或一個迫擊砲會爆炸與否，我們都無法事先預測，所以，從預言家的觀點來看，就因為這個不確定性讓我們的生活多采多姿，而生活的不確定性讓我們可以努力挑戰更美好的生活。在現實生活中，(幾乎) 每件事情都不確定，所以，當我們做決策時，我們就把這不確定性考慮進去。但是無論一個人怎麼去計算機率，一件事發生的機率如果接近 1，表示這件事非常可能發生，而機率如果接近 0 的話，表示這件事不太可能發生。

為了瞭解機率的理論，我們首先看看 (統計) 實驗的基本觀念，從實驗當中，我們可以在每一事情上看到不確定性。

I. 實驗和樣本空間

實驗 (experiment)：一個結果不確定的機率過程稱為實驗 (或統計實驗)。例如：

- 例題 1：丟擲一個公平的銅板，直到丟擲結束前其結果是未知的，因此，擲一個銅板是一個實驗。
- 例題 2：擲一顆公平的骰子並觀察出現的點數，在骰子停止前，我們並不會知道出現的點數為何，因此，擲骰子也是一種實驗。

給定一個特別實驗，實驗所可能發生結果的集合稱為樣本空間 (sample space)，通常以符號 S 表示，樣本空間的一個元素稱為樣本點 (sample point)。

上述例題 1 和例題 2，其樣本空間分別為

- 例題 1：S={ 正面 , 反面 }
- 例題 2：$S=\{1, 2, 3, 4, 5, 6\}$

II. 事件和事件的基本運算方法

事件：任何一個樣本空間 S 的子集合 (subset) 稱為事件，通常事件以 A, B, C 等符號表示之。

一個事件如果沒有任何元素，稱為空的或虛無 (null) 事件，以符號 ϕ 表示之。

- 例題 1 (續)：S={ 正面 , 反面 }，令 A={ 正面 }，則 A 是一個事件，令 B 是一個結果不是正面也不是反面的事件，則 $B = \phi$。
- 例題 2 (續)：$S=\{1, 2, 3, 4, 5, 6\}$，令 $A=\{1, 3, 5\}$ ={ 出現點數為奇數 }，則 A 是一個事件。

接下來我們定義事件的一些基本運算符號：

1. 聯集 (union)：對兩事件 A 和 B， $A \cup B$ (稱為 A 和 B 的聯集) 表示不是屬於 A，就是屬於 B，為兩者元素的集合。
2. 交集 (intersection)：對兩事件 A 和 B， $A \cap B$ (稱為 A 和 B 的交集) 表示元素同時屬於 A 和 B 的集合，如果 $A \cap B = \phi$(表示 A 和 B 之間沒有共同點)，那麼兩事件 A 和 B 被稱為無交集 (disjoint) 或互斥 (mutually exclusive)。
3. 事件的補集 (complement)：對任一事件 A，A^C(稱為 A 的補集) 表示不是屬於事件 A 元素的集合。

III. 機率及機率法則

III-1. 機率

給定一個實驗及其樣本空間，我們可以找出任何事件 A 的機率，以符號 $P(A)$ 表示，$P(A)$ 可以以下列兩種方式求之。

1. 機率的標準定義：假設我們有一實驗其樣本空間 S 內之元素是有限的，每個元素發生的可能性相同，那麼一個事件 A 的機率 $P(A)$ 被定義為

$$P(A) = \text{樣本 } A \text{ 裡的元素個數} / \text{樣本 } S \text{ 空間裡的元素個數}$$

- 例題 1 (續)：假設擲一個公平的銅板一次，則樣本空間 S={ 正面 , 反面 }，令 A={ 正面 } 是一個事件，所以 $P(A)$ = 正面的機率 = 0.5。
- 例題 2 (續)：丟擲一個公平的骰子並計算出現點數小於 3 的機率。樣本空間 S={1, 2, 3, 4, 5, 6}，因為骰子是公平的，所以所有可能結果發生的可能性相同，定義事件 A={1, 2}，那麼 $P(A)$ = 觀察到 1 或 2 的機率 =0.3333。

2. 機率就是相對次數：假設一個實驗重複做了 n 次，令 A 表示一個在實驗中可能發生的事件，$N(A)$ 表示在 n 次實驗中事件 A 發生的次數，而 $f(A) = N(A)/n$ 表示事件 A 的相對次數，當 n 愈來愈大時，$N(A)$ 所趨近的值可以被解釋為在單一實驗中事件 A 發生的機率，換句話說，假如相同的實驗被重複非常多次，那麼事件發生的相對次數就近似於事件發生的機率。

III-2. 機率的基本法則

根據上述機率的解釋，下列機率法則可以應用在任何實驗

1. 對任何事件 A，$0 < P(A) < 1$。
2. $P(S)$ =1 及 $P(\phi)$ =0。

3. 對兩無交集的事件 A 和 B，$P(A \cup B) = P(A) + P(B)$。通常，對任兩事件 A 和 B，$P(A \cup B) = P(A) + P(B) - P(A \cap B)$。

4. 對任何事件 A，$P(A^C) = 1 - P(A)$。

III-3. 條件機率

在前面的章節中，我們已知道如何求得實驗中某一事件發生的機率，但這個機率可能在某個事件已經發生的情況下而改變，我們用下列例題來解釋這種情形。

- 例題 3：考慮從一副 52 張牌中，隨機抽取一張牌，則樣本空間為這 52 張牌，令 A 為抽到 K 的事件，那麼抽到 K 的機率為何？答案是 $P(A)$ = 4/52 = 1/13。如果我們在抽這張牌時，已知道這張牌不是 Q，那麼這時抽到 K 的機率為何？換句話說，這個額外知道不是 Q 的資訊，是否會影響 $P(A)$ 的機率？

  答：這個問題的答案是肯定的，其機率為

  P (當我們知道不是 Q 時，抽到的牌是 K) = 4/48 = 1/12。

為什麼上述兩個機率會不同呢？我們注意到我們已經知道牌不是 Q，定義事件 B 為抽到的牌不是 Q 的事件，B 包含所有 48 張不是 Q 的牌，因此 $P(B)$ = 48/52。

當我們知道牌不是 Q 時，抽到 K 的情形，以機率的名詞來說，可以稱為“給定 B 已經發生的情況下，A 事件發生”，或以符號“$A|B$”表示，簡稱 A 給定 B。“給定 B 已經發生的情況下，A 事件發生”的機率，稱為“A 給定 B 的條件機率 (conditional probability)”，其定義為 $P(A \mid B) = P(A \cap B) / P(B)$。

在例題 3 中，$A \cap B$ = { 所有 K 的牌 } ∩ { 所有非 Q 的牌 }={ 所有 K 的牌 }

也就說 $P(A \cap B) = 4/52 = 1/13$，因此

$$P(A \mid B) = P(A \cap B) / P(B) = (1/13)/(12/13)$$

兩事件 A 和 B 是互相獨立 (independent) 的，如 $P(A|B) = P(A)$；或 $P(B|A) = P(B)$；或 $P(A \cap B) = P(A)P(B)$。上述三個條件是全等的，也就是說任何一個條件成立，另二個條件也成立，根據上述條件，給定事件 B，A 事件的條件機率與無條件的事件 A 之機率相同，這意謂 B 事件的發生不會影響 A 事件的發生，這就是所謂事件 A 和 B 是互相獨立的。而 $P(A \cap B) = P(A)P(B)$ 則被廣泛地用來證明兩個事件的獨立性。如兩事件不獨立，則我們稱兩事件是相依的 (dependent)。

IV. 隨機變數之期望值及變異數

隨機變數是把實驗結果以數字來表示，也就是把樣本空間對應到實數上的一個函數，通常以符號 X、Y 或 Z 來表示。而隨機變數 X 之期望值 (Expected Value) 或平均值的定義為：

$$\mu = E(X) = \sum_{i=1}^{k} x_i P(X = x_i)\text{。}$$

另外，隨機變數 X 之變異數定義為：

$$\sigma^2 = Var(X) = V(X) = \sum_{i=1}^{k} (x_i - \mu)^2 P(X = x_i) = E(X^2) - \mu^2\text{。}$$

有關隨機變數 X 之期望值、變異數的重要性質有：

(a) $E(a) = a$，如 a 為常數。

(b) $E(aX) = aE(X)$，如 a 為常數。

(c) $E(aX+b) = aE(X)+b$，如 a、b 為常數。

(d) $Var(a) = 0$，如 a 為常數。

(e) $Var(aX) = a^2 Var(X)$，如 a 為常數。

(f) $Var(aX+b) = a^2 Var(X)$，如 a、b 為常數。

如 X 和 Y 為兩隨機變數，a 和 b 為常數，則期望值、變異數的重要性質有：

(a) $E(aX+bY) = aE(X)+bE(Y)$。

(b) $Var(aX \pm bY) = a^2Var(X) + b^2Var(Y) \pm 2abCov(X,Y)$

$= a^2Var(X) + b^2Var(Y)$，如兩隨機變數 X、Y 是互相獨立。

演練題 (Q & A)

一、是非題

1 ：已知 E 與 F 是兩事件，若 $E \subset F$，則 $P(E^C) + P(F) \le 1$。

Ans X

2 ：對任意互斥事件 E 與 F，則 $P(E \cap F) = P(E)P(F)$。

Ans X

3 ：若 A 與 B 兩事件為互斥 (mutually exclusive) 事件，則 $P(A \cup B) = 1$。

Ans X

4 ：A 與 B 兩事件不可能同時發生，則稱此兩事件為互斥事件 (mutually exclusive)。

Ans O

5 ：若 X 是一隨機變數，則 $E(X^2) = (E(X))^2$。

Ans X

6 ：若 X 與 Y 是互相獨立的隨機變數，則 $Var(X+Y) = Var(X-Y)$。

Ans O

7 ：若 X 與 Y 皆是隨機變數，則 $E(XY) = E(X)E(Y)$。

Ans X

8 ：若 X 與 Y 皆是隨機變數，則對所有 $a \in \mathrm{R}$，$Cov(aX, Y) = Cov(aY, X)$。

Ans O

9 ：若隨機變數 X 與 Y 互相獨立，則 X 與 $2Y+1$ 互相獨立。

Ans O

10 ：若 X 與 Y 皆是隨機變數，則 $P(A \cap B) = 0$。

Ans X

11 ：若 A 與 B 兩事件為獨立 (statistically independent) 事件，則 $P(A \cap B) = 0$。

Ans X

二、選擇題

1 ：連續丟擲一顆骰子，直到點數 3 出現為止，令 E_n 表總共丟擲的次數，求 $P(E_1 \cup E_2)$值。

A：2/36

B：5/36

C：6/36

D：11/36

Ans D

2 ：事件 $(E \cup F) \cap (E \cup F^c)$ 的簡單表示式為何？

A：E

B：ϕ

C：$E \cap F$

D：$E \cap F^c$

Ans A

3 ：A, B 兩個箱子，A 箱子內含編號 1, 3 的兩個球，B 箱子內含編號 2, 4, 6 的三個球，丟擲一枚公平硬幣，隨機選定一箱子，再由此箱子隨機抽出一球，試問抽到編號 6 的球之機率為何？

A：1/3

B：5/6

C：1/6

D：1/4

Ans C

4 ：有兩個電子零件，已知第一個零件正常運作的機率為 90%，假設第一個零件正常運作下，第二個零件是故障的機率是 10%，且當第一個零件故障下，第二個零件是故障的機率是 30%。試問只有一個零件是正常運作的機率為何？

A：0.16

B：0.97

C：0.79

D：0.0063

Ans A

5 ：已知二獨立事件的發生機率分別為 0.2 與 0.4，試問此二事件至少有一事件發生的機率為何？

A：0.44

B：0.52

C：0.60

D：0.92

Ans B

6 ：若兩事件 A 與 B 機率分別為 $P(A) = 0.8$ 與 $P(B) = 0.6$，則 $P(A \cap B)$ 的最小值為何？

A：0.4

B：0.6

C：0.2

D：0.8

Ans A

7 ：令 X 表定義在 -1, 0, 1 的隨機變數，且 $P(X=-1)=0.3$, $P(X=0)=0.5$, $P(X=1)=0.2$，求 $E(X^2)$ 的值。

A：−0.1

B：0.01

C：0.38

D：0.5

Ans D

8 ：丟擲兩顆公平骰子一次，令隨機變數 X 表其出現的點數和，求 $P(X=6)$ 的值。

A：3/36

B：4/36

C：5/36

D：6/36

Ans C

9 ：丟擲兩顆公平骰子一次，令隨機變數 Y 表其出現的點數之乘積，求 $E(Y)$。

A：21/6

B：21/36

C：21/9

D：49/4

Ans D

10 ：丟擲一公平骰子，若 X 表其出現的點數，求 $Var(X)$。

A：49/4

B：70/6

C：91/6

D：35/12

Ans D

11 ：採購商購買 10 個為一組的電腦零件。他的採購原則是，隨機檢驗一組內 4 個零件品，若此 4 個零件皆正常，則採購此組電腦零件。若 30% 的電腦零件組具有 3 個故障品，而其他 70% 的電腦零件組僅有一個故障品，試問採購商購買此電腦零件組的機率為何？

A：47/100

B：53/100

C：50/63

C：13/63

Ans A

12 ：已知 X 與 Y 是互相獨立且具有相同分配的隨機變數，且 $E(X) = 1$，$Var(X) = 4$，求 $Var(3X - 2Y)$。

A：4

B：20

C：52

D：100

Ans C

13 ：若兩事件 A 與 B 為周延 (collectively exhaustive)，則 $P(A \cap B) =$ ________ 。

A：$P(B) + P(A) - 1$

B：$P(A)P(B)$

C：$P(B) + P(A)$

D：$P(A) / P(B)$

Ans A

14 ：已知 $E(X^2) = 4$、$E(Y^2) = 3$、$E(XY) = 1$，求 $E((X - Y^2))$。

A：1

B：5

C：6

D：9

Ans B

15 ：若兩事件 A 與 B 機率分別為 $P(A) = 0.4$ 與 $P(B) = 0.5$ 且 $P(A \cap B) = 0.2$，則 $P(B \mid A) =$ ______。

A：0.4

B：0.5

C：0.2

D：0.7

Ans B

16 ：若兩事件 A 與 B 機率分別為 $P(A) = 0.4$ 與 $P(B) = 0.5$ 且 $P(A \mid B) = 0.2$，則 $P(B \mid A)$ =______。

A：0.4

B：0.5

C：0.25

D：0.1

Ans C

17 ：已知 X、Y 是互相獨立隨機變數，且 $X \sim B(10, 1/2)$，$Y \sim B(10, 1/4)$，求 $Var(X - Y)$ 值。

A：10/16

B：60/16

C：70/16

D：80/16

Ans C

18 ：已知袋子內有 4 顆白球與 7 顆紅球，隨機取出 1 球，

令$X = \begin{cases} 2, \text{取出的是白球} \\ 1, \text{取出的是紅球} \end{cases}$，求 $E(X)$。

A：15/11

B：16/12

C：46/33

D：13/33

Ans A

19 ：若兩事件 A 與 B 機率分別為 $P(A) = 0.4$ 與 $P(B) = 0.5$，則條件機率 $P(A \mid B)$ 的最大值為何？

A：0.4

B：0.2

C：0.5

D：0.8

Ans D

20 ：若兩事件 A 與 B 機率分別為 $P(A)=0.4$ 與 $P(B)=0.6$，則 $P(A\cup B)$ 的最大值為何？

A：0.4

B：0.6

C：1.0

D：0.8

Ans C

三、複選題

1 ：若事件 E 與 F 互相獨立 (statistically independent)，則下列何者正確。

A： $P(E\cap F)=P(E)P(F)$ 。

B： $P(E\cap F^c)=P(E)P(F^c)$ 。

C： $P(E\cup F)=P(E)+P(F)$ 。

D： $P(E\cap F)=0$ 。

E： $P(E\,|\,F)=P(E)$ 。

Ans A, B, E

2 ：若事件 E 與 F 互斥 (mutually exclusive)，則下列何者正確。

A： $P(E\cap F)=P(E)P(F)$ 。

B： $P(E\cap F^c)=P(E)P(F^c)$ 。

C： $P(E\cup F)=P(E)+P(F)$ 。

D： $P(E\cap F)=0$ 。

E： $P(E\,|\,F)=0$ 。

Ans C, D, E

單　元

4

機率分配模型

機率的型態分為兩類，離散型機率分配及連續型機率分配。在本單元我們將簡介幾個常用的機率分配，包括離散型機率分配中的二項分配 (Binomial Distribution)、波氏分配 (Poisson Distribution) 及連續型機率分配中的常態分配 (Normal Distribution) 、均勻分配、指數分配。

I. 二項分配

介紹二項 (Binomial) 隨機變數之前，一定要先介紹貝努利 (Bernoulli) 隨機變數。

I-1. 貝努利機率密度函數定義

當一試驗 (Trial)，其結果可被區分為“成功 (x = 1)”或“失敗 (x = 0)”兩類情況之下，則隨機變數 X 代表一次試驗中成功的次數，其機率密度如下所示

$$P(X=x)=\begin{cases}1-p, & \text{if } x=0\\ p, & \text{if } x=1\end{cases}$$

其中試驗成功的機率為 p 且 $0 \le p \le 1$，則稱 X 為貝努利隨機變數；簡記為 X~Bernoulli(p)。

註

Bernoulli(1654-1705) 為瑞士數學家。

I-2. 二項分配機率密度函數與累積分佈函數定義

二項分配機率密度函數的定義如下：假設進行 n 次獨立貝努利試驗，每次試驗成功的機率均為 p，則 n 次試驗中成功的次數 X 的機率密度函數為

$$P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}, x=0, 1, 2\cdots, n$$

且隨機變數 X 稱為二項隨機變數 (Binomial Random Variable)，簡記為 $X\sim B(n, p)$。

二項的累積分佈函數 (Cumulative Distribution Function) 的定義如下：

$$P(X \le x) = \sum_{k=0}^{x}\binom{n}{k}p^k(1-p)^{n-k},\ x = 0, 1, 2, \cdots, n$$

I-3. 二項分配基本性質

1. 貝努利機率密度函數其實是二項分配的特例；即 Bernoulli(p) ≡ $B(1, p)$。
2. 若 $X_i \overset{iid}{\sim}$ Bernoulli(p) ≡ $B(1, p)$，$i = 1, 2\cdots, n$，則 $Y = \sum_{i=1}^{n} X_i \sim B(n, p)$。
3. 若 $X\sim B(n, p)$，則 $E(X) = np$，$Var(X) = np(1-p)$。
4. 若 $X\sim B(n, p)$，$Y\sim B(m, p)$ 且 X 和 Y 互相獨立，則 $X+Y\sim B(n+m, p)$。

I-4. 二項分配機率的求取

範例 4.1 投擲均勻銅板 3 次，假設其出現的結果皆互相獨立，試求出現正面個數的 (1) 機率密度函數、(2) 累積機率密度函數表。

Ans 設 X：出現正面個數 且 $X\sim B(3, 0.5)$

(1) $P(X=0) = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \frac{1}{8}$，$P(X=1) = \binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 = \frac{3}{8}$

$P(X=2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8}$，$P(X=3) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8}$

所以，X 的機率密度函數 *p.d.f.* 為

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

(2) $P(X \le 0) = P(X = 0) = 1/8$

$P(X \le 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 4/8$

$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 1/8 + 3/8 + 3/8 = 7/8$

$P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1/8 + 3/8 + 3/8 + 1/8 = 1$

所以，X 的累積機率密度函數 *c.d.f.* 為

x	0	1	2	3
$P(X \le x)$	1/8	4/8	7/8	1

II. 波氏分配

定義 設隨機變數 X 之機率分配為

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & , x = 0, 1, 2, \ldots; \lambda > 0 \\ 0 & ,\text{其他} \end{cases}$$

則我們稱為 X 為具參數 $\lambda > 0$ 之波氏分配，記為 X~Poisson(λ)，且 $E(X) = \lambda$，$Var(X) = \lambda$。

III. 常態分配

常態分配幾乎是所有機率分配中最重要的一個機率分配，因為在實際生活中有太多的變數都是服從或近似於常態分配。一般來說，若我們用一條特定平滑的對稱曲線來近似一個類似鐘型的機率直方圖，那這條平滑曲線就被稱為常態曲線 (Normal Curve)。在 1733 年時 Abraham DeMoiver 導出了常態曲線的數學方程

式，而該方程式已成為近代科學研究領域中一個非常重要的理論基礎。另外 Karl Friedrich Gauss 亦在其研究中導出這條方程式，有時候為了紀念他，常態曲線也會被稱為高斯曲線 (Gaussian Curve)。

III-1. 常態分配機率密度函數的定義

下面是常態分配所對應的機率密度函數之定義。一個服從平均數為 μ，標準差為 σ 常態分配 $N(\mu, \sigma)$ 的隨機變數，其機率密度函數如下：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},-\infty<x<\infty;-\infty<\mu<\infty;\sigma>0$$

其中 $e = 2.71828\cdots$且 $\pi = 3.14159\cdots$。

III-2. 常態分配基本性質

接著我們來看看常態分配所具備的性質有那些：

1. 常態曲線是一個對稱的鐘型曲線。
2. 常態曲線下的總面積為 1(此為機率公理，任何機率密度函數皆滿足此一條件)。
3. 常態隨機變數 X 的值是介於 $-\infty$ 和 ∞ 之間，且常態曲線在兩端會趨近於平行軸 X，但不會和 X 軸相交或穿越 X 軸。
4. 常態曲線的兩個特徵 (參數)

❶ 出現中心高峰的位置，通常以平均數 μ 來表示。

❷ 資料離散度的情形，通常以標準差 σ 來表示。

隨著不同的特徵 (μ 和 σ)，畫出的常態曲線就會有所不同，下面是可能遇到的一些情況：

■ 情況一

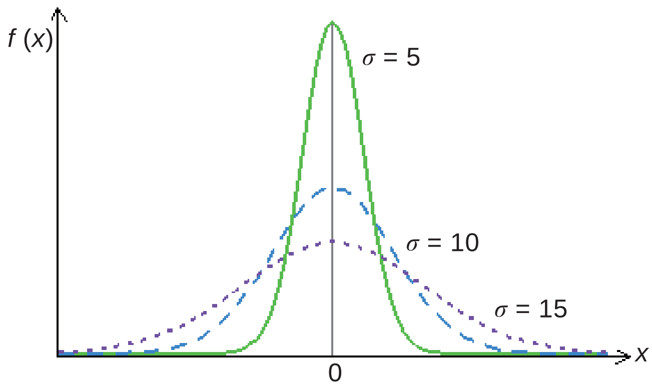


圖 4.1：μ = 0，但 σ 值不同時的常態曲線圖

■ 情況二

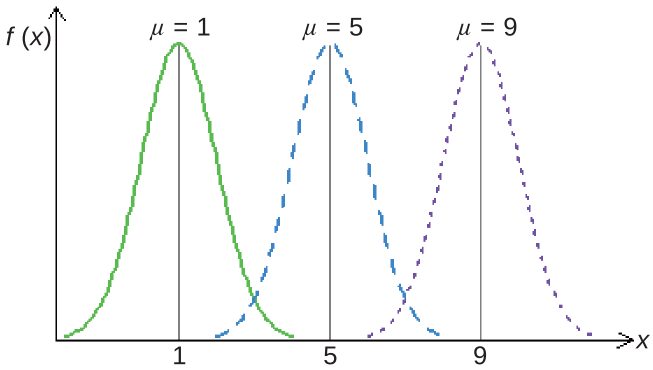


圖 4.2：σ = 1，但 μ 值不同時的常態曲線圖

■ 情況三

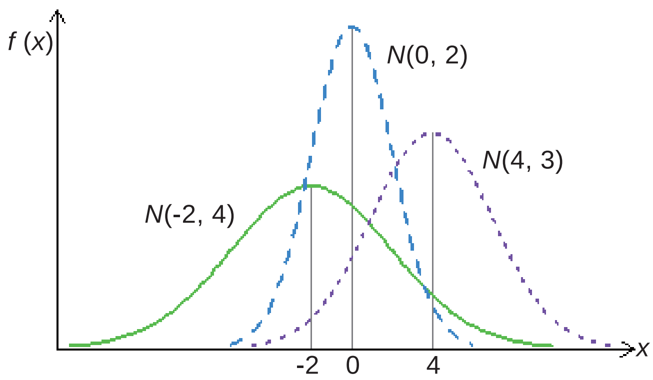


圖 4.3： μ 和 σ 皆不相同時的常態曲線圖

經過圖 4.1 至圖 4.3 的圖示說明，雖然圖 4.4 並未標示平均數及變異數的資訊，但應該很容易可以判斷圖 4.4 中常態曲線 2 的期望值最大，而常態曲線 3 的變異數最大。

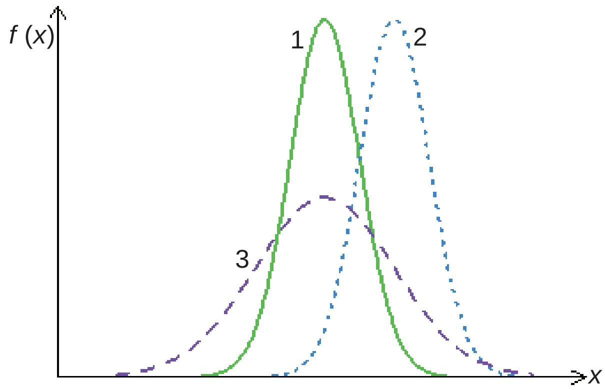


圖 4.4： 不同常態曲線平均數與變異數的比較

註

在連續型機率分配中，面積代表機率，所以使用圖形來解說機率求取的概念是最佳方式之一；而不同參數底下，繪製出來的機率密度函數圖形的形狀也會不同，所以如果在書中的例子做精確圖形的呈現，則會佔掉太多的篇幅；為了兼顧圖示說明的方便性及節省篇幅的考量，我們採限制圖形高度的方式來繪圖，雖然有點失真，但也不失為一個兼顧學習效果及節省篇幅考量的折衷辦法。

III-3. 常態分配機率的求取

我們知道連續型隨機變數 X 其機率密度函數曲線以下，在 X 軸某特定範圍之上所包圍的面積，即代表 X 在該特定範圍的機率，即如圖 4.5 所示：

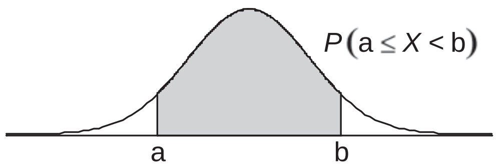


圖 4.5：X 在該特定範圍的機率

那要如何求得這個面積 (機率值) 的大小呢？一般而言，積分是一個最常用來求不規則形面積的技巧，但因為常態分配的機率密度函數公式過於複雜，導致無法直接使用積分求出其任意範圍所對應的機率值，所以傳統上多採用查詢統計機率表的方式求取機率。不過經由先前的討論可以知道，常態分配是一個大家族，每一條常態曲線都有它自己獨特的特徵 (μ 和 σ)，那是不是代表我們須準備無數多個常態分配機率表，才能夠協助我們來找出各種可能所對應的機率值呢？所幸不是，事實上我們只需要透過標準常態分配表就可以求出對應的常態曲線下的面積了。

假設 X 服從 $N(\mu, \sigma)$，透過如下標準化的程序

$$Z = \frac{X - \mu}{\sigma}$$

則 Z 會服從 $N(0, 1)$，其中 $N(0, 1)$ 即是所謂的標準常態分配。

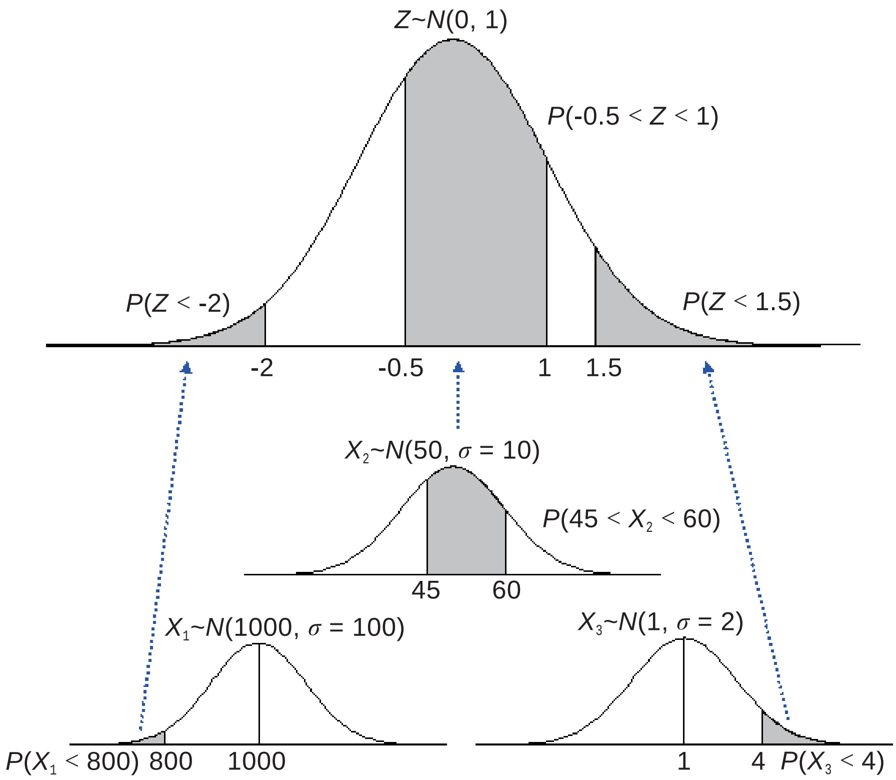


圖 4.6：一般常態分配機率求取與標準常態分配 (Z) 之間關係的示意圖

上面的圖示，呈現了一般常態分配機率求取與標準常態分配 (Z) 之間的關係。

III-4. 常態分配切點的求取

在許多統計問題中，我們亦有可能會遇到另外一種狀況，就是在給定某機率值的條件下如何找到隨機變數可以滿足以上條件的位置點，即所謂的切點 (Cutoff Point)。比如說假設台灣 10 歲到 12 歲青少年身高服從常態分配 $N(\mu = 160, \sigma = 5)$，那台灣 10 歲到 12 歲青少年前 10% 身高是幾公分以上呢？

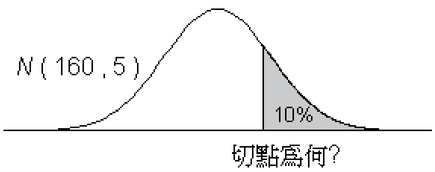
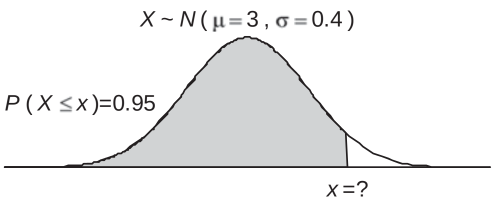


圖 4.7： 常態分配的切點

常態分配對應某範圍的切點時，是無法直接透過查表求得的；傳統上的做法是先找出對應的標準常態分配的切點，再轉換找出對應的切點，而轉換的公式為 $x = \mu + z \times \sigma$。

範例 4.2 在常態分配 $N(\mu = 3, \sigma = 0.4)$ 中，求滿足左尾機率 $P(X \leq x) = 0.95$ 的切點 x。



Ans $P(X \leq x) = P(\frac{X-3}{0.4} \leq \frac{x-3}{0.4}) = P(Z \leq \frac{x-3}{0.4}) = 0.95$

由查表得知，$P(Z \leq 1.645) = 0.95 \Rightarrow \frac{x-3}{0.4} = 1.645$

所以，切點 $x = 3 + 1.645 \times 0.4 = 3.658$

IV. 均勻分配

定義 設隨機變數 X 之機率分配為

$$f(x)=\begin{cases}1/(b-a) & ,\ a<x<b\\ 0 & ,\text{其他}\end{cases}$$

則我們稱 X 為具參數 $a,\ b$ 之均勻分配，記為 $X\sim(a,\ b)$，且 $E(X)=(a+b)/2$，$Var(X)=(b-a)^2/12$。

V. 指數分配

定義 設隨機變數 X 之機率分配為

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,\ 0\le x\\ 0 & ,\text{其他}\end{cases}$$

則我們稱 X 為具參數 $\lambda>0$ 之指數分配，記為 $X\sim Exp(\lambda)$，且 $E(X)=1/\lambda$，$Var(X)=1/\lambda^2$。

演練題 (Q & A)

一、是非題

1 ：$P(X=0)=0.1$，$P(X=1)=-0.2$，$P(X=2)=0.3$，$P(X=4)=0.3$，$P(X=5)=0.5$，滿足離散機率分配之條件。

Ans X

2 ：令隨機變數 X 的機率分配如下：

x	-1	1	3	6
$P(X = x)$	0.4	0.2	0.2	0.2

令 $Y = X^2$，則 $P(Y = 1) = 0.6$。

Ans O

3 ：Austin 參加一個測驗，共有 25 題五選一單選題，若他全部用猜的，則他猜中題數的變異數為 4。

Ans O

4 ：某銀行在每一小時內的顧客出現人次為 $\mu = 6$ 的 Poisson 波氏分配。在 20 分鐘內來顧客人數的變異數與期望值相等。

Ans O

5 ：已知班上 50 名同學中有 20 名女生 30 名男生。現在由此班隨機抽出 10 名同學，則抽中男生人數的期望值大於其變異數。

Ans O

6 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則變異數 $Var(X) = 4$。

Ans X

7 ：令 X 為平均數 $\mu = 4$ 之指數分配隨機變數，則變異數 $Var(X) = 4$。

Ans X

8 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則四分位距 (IQR) 為 3.5。

Ans O

9 ：令 Z 為標準常態分配，則機率 $P(Z > 1) = 0.1587$。

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

Ans O

10：令 X 為平均數 $\mu = 4$ 之指數分配隨機變數，則 X 的四分位距 (IQR) 為 4。

Ans X

二、選擇題

1：若 X 表示 n 次獨立相同的貝努利實驗中成功的總次數，則 X 的機率分配為何種分配？

A：常態分配

B：超幾何分配

C：幾何分配

D：二項分配

Ans D

2 ：X 為幾何分配，其成功率 $p = 0.1$，則 $P(X = 5)$ 即是實驗在第五次得到成功的機率為何？

A：0.9×0.1

B：$0.9 \times 4 \times 0.1$

C：$0.9^4 \times 0.1$

D：0.1×5

Ans C

3 ：若 X 為幾何分配，$x = 1, 2, \cdots$，其中成功率 $p = 0.1$，則 X 之期望值 $E(X)$ 為何？

A：10

B：5

C：2

D：3.5

Ans A

4 ：若 X 為幾何分配，$x = 1, 2, \cdots$，其中成功率 $p = 0.1$，則 X 的變異數 $Var(X)$ 為何？

A：10

B：90

C：50

D：45

Ans B

5 ：若 X 為不連續均等分配 U(9) 之隨機變數，$x = 1, 2, \cdots, 9$，則機率 $P(X = 10)$ 為何？

A：0.1

B：1/9

C：0.9

D：0

Ans D

6 ：若 X 為不連續均等分配 U(9) 之隨機變數，$x = 1, 2, \cdots, 9$，則 X 之期望值 $E(X)$ 為何？

A：5

B：9

C：10

D：2

Ans A

7 ：若 X 為不連續均等分配 U(9) 之隨機變數，$x = 1, 2, \cdots, 9$，則 X 之變異數 $Var(X)$ 為何？

A：7.500

B：4.500

C：6.667

D：9.000

Ans C

8 ：令隨機變數 X 的機率函數為 $P(X = x) = x/10$，其中 $x = 1, 2, 3, 4$。則 X 的期望值 $E(X) =$ ？

A：2.5

B：2

C：3

D：4

Ans C

9 ：令隨機變數 X 的機率函數為 $P(X = x) = x/10$，其中 $x = 1, 2, 3, 4$。則 X 的變異數 $Var(X) =$ ？

A：1

B：2

C：2.5

D：3

Ans A

10 ：令 X 為一個離散型均等分配 U(7) 之隨機變數，機率函數 $p(x) = 1/7$，$x = 1, 2, \cdots, 7$，此分配資料在平均數加減一倍標準差 (包含兩個端點) $[\mu-\sigma, \mu+\sigma]$ 的機率為何？

A：0.5

B：0.83

C：0.714

D：0.62

Ans C

11 ：令 X 為一個離散型均等分配 U(7) 之隨機變數，機率函數 $p(x) = 1/7$，$x = 1, 2, \cdots, 7$，此分配資料在平均數加減 2 倍標準差 (包含兩個端點) $[\mu-2\sigma, \mu+2\sigma]$ 的機率為何？

A：0.86

B：0.95

C：0.997

D：1.00

Ans D

12 ：購買刮刮樂一張花費 10 元，若刮中可兌換獎金 1000 元。假設刮刮樂中獎機率為千分之一，試問購買此刮刮樂的期望值為何？

A：0

B：-1

C：-9

D：-5

Ans C

13 ：隨機變數 X 的期望值與變異數分別為 $E(X) = 2$，$Var(X) = 5$。令 $Y = 2X+1$，則 Y 的期望值 $E(Y) =$ ？

A：2

B：3

C：5

D：10

Ans C

14 ：隨機變數 X 的期望值與變異數分別為 $E(X) = 2$，$Var(X) = 5$。令 $Y = 2X+1$，則 Y 的變異數 $Var(Y) =$ ？

A：5

B：10

C：20

D：25

Ans C

15 ：已知班上 50 名同學中有 20 名女生 30 名男生。現在由班上隨機抽出一名同學，若該名學生為女性則令隨機變數 $X = 0$，若為男性則令 $X = 1$。則 X 的變異數 $Var(X) =$ ？

A：0.4

B：0.6

1335-CC1402-Ch4-Ch6.indd 16 2014/10/22 下午 12:33:15

C：20

D：0.24

Ans D

16 ：Austin 參加一個測驗，共有 25 題五選一單選題，若他全部用猜的，則他猜中 10 題的機率為何？

A：$C^{25}_{10}0.2^{10}\times0.8^{15}$

B：$C^{25}_{10}0.2\times0.8$

C：$0.2^{10}\times0.8^{15}$

D：5^{10}

Ans A

17 ：Austin 參加一個測驗，共有 25 題五選一單選題，若他全部用猜的，則他猜中題數的期望值為何？

A：6.25

B：10

C：5

D：12.5

Ans C

18 ：Austin 參加一個測驗，共有 25 題五選一單選題，若他全部用猜的，則他猜中題數的變異數為何？

A：5

B：4

C：10

D：12.5

Ans B

19 ：某銀行在每一小時內顧客出現的人次為 μ = 6 的 Poisson 波氏分配。試問在 20 分鐘內有 2 位顧客光臨的機率為何？

A：$C_2^6 0.1^2 \times 0.9^4$

B：$2e^{-2}$

C：e^{-2}

D：0.333

Ans B

20 ：某銀行在每一小時內的顧客出現的人次為 μ = 6 的 Poisson 波氏分配。試問在 20 分鐘內來顧客人數的期望值為何？

A：3

B：6

C：2

D：4

Ans C

21 ：某銀行在每一小時內顧客出現的人次為 μ = 6 的 Poisson 波氏分配。試問在 20 分鐘內來顧客人數的變異數為何？

A：3

B：1

C：2

D：4

Ans C

22 ：已知班上 50 名同學中有 20 名女生 30 名男生。現在由班上隨機抽出 10 名同學，則抽中 4 名女生 6 名男生的機率為何？

A：$C_6^{30} C_4^{20} / C_{10}^{50}$

B：$C_6^{30} C_4^{20}$

C：$0.6^6 \times 0.4^4$

D：$C_6^{10} 0.6^6 \times 0.4^4$

Ans A

23 ：已知班上 50 名同學中有 20 名女生 30 名男生。現在由班上隨機抽出 10 名同學，則抽中男生人數的期望值為何？

A：4

B：6

C：10

D：5

Ans B

24 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則機率 $p(X = 3.5) =$ ？

A：1/7

B：1/2

C：1/3.5

D：0

Ans D

25 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則期望值 $E(X) =$ ？

A：7

B：4

C：3.5

D：5

Ans D

26 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則變異數 $Var(X) =$ ？

A：4

B：3

C：49/12

D：3.5

Ans C

27 ：令 X 為指數分配之隨機變數且變異數 $Var(X) = 4$，則平均數 $E(X) =$ ？

A：4

B：16

C：2

D：8

Ans C

28 ：令 X 為平均數 $E(X) = 3$ 之指數分配隨機變數，則變異數 $Var(X) =$ ？

A：3

B：6

C：9

D：1

Ans C

29 ：令 X 為平均數 $E(X) = 3$ 之指數分配隨機變數，則機率 $P(X > 6) =$ ？

A：e^{-2}

B：$1-e^{-1}$

C：e^{-1}

D：e^{-6}

Ans A

30 ：令 X 為平均數 $E(X) = 4$ 之指數分配隨機變數，則條件機率 $P(X > 6 \mid X > 2) =$ ？

A：$P(X > 6)$

B：$P(X > 2)$

C：$P(X > 4)$

D：$P(X > 8)$

Ans C

31：標準常態分配的四分位距 (*IQR*) 約為多少？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

A：1

B：1.5

C：1.34

D：2

Ans C

32 ：標準常態分配具有以下何種性質？

A：平均數為 0 且標準差為 1

B：平均數為 1 且標準差為 0

C：中位數為 1

D：無法使用於離散分配的估計

Ans A

33 ：以下關於常態分配的敘述，何者為不正確？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

A：平均數、中位數與眾數三者相等

B：大約有 2/3 的資料會落在平均數加減一倍標準差內

C：是一種離散分配

D：大約有 95% 的資料會落在平均數加減二倍標準差內

Ans D

34 ：假設某資料分配近似於常態分配，我們可以得到以下何結果

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：大約有 68% 的資料會落在平均數加減一倍標準差內

B：大約有 4/5 的資料會落在平均數加減 1.282 倍標準差內

C：大約有 9/10 的資料會落在平均數加減 1.645 倍標準差內

D：所列三項皆是

Ans D

35 ：令 X 為具平均數 10 且標準差 5 的常態分配之隨機變數，則機率 $P(-5 < X < 30)$ = ？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：0.9987

B：0.4772

C：0.3413

D：0.8021

Ans A

36 ：某銀行上午營業時段內來客間隔時間是依照指數分配，其平均間隔時間為 4 分鐘，試問在 8 分鐘內均無來客的機率為何？

A：e^{-8}

B：e^{-1}

C：e^{-4}

D：e^{-2}

Ans D

37 ：若成人的智商 (*IQ*) 資料結構是依照平均數為 100 標準差為 20 的常態分配，試問最聰明的前 5% 成人的智商約為多少以上？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：120

B：118

C：133

D：130

Ans C

38 ：以下何種二項分配 $B(n = 40, p)$ 的單次成功機率 p，如使用常態近似二項機率結果會產生較大的誤差。

A：0.9

B：0.2

C：0.65

D：0.3

Ans A

39 ：已知隨機變數 X 的分配函數 $F(x)=\begin{cases}0, x<0\\ 1/2, 0\le x<1\\ 2/3, 1\le x<2\\ 1, 2\le x\end{cases}$. 求 $P(X=1)$ 的值。

A：1/6

B：1/3

C：1/2

D：2/3

Ans A

三、複選題

1 ：令 X 為連續型均等分配 U(0, 7) 之隨機變數，則下列哪些為正確？

A：機率 $P(X > 3.5) = 1/7$

B：機率 $P(X > 3.5) = 1/2$

C：期望值 $E(X) = 3.5$

D：變異數 $Var(X) = 4$

E：平均數小於標準差

Ans B, C

2 ：以下關於常態分配的敘述那些為正確。

A：平均數、中位數與眾數三者相等

B：Midrange=Midquartile

C：是一種離散分配

D：大約有 95% 的資料會落在平均數加減二倍標準差內

E：不可以被使用來估計離散分配機率

Ans A, B, D

3 ：令 X 為平均數 $E(X) = 4$ 之指數分配隨機變數，則

A：X 的變異數 $Var(X) = 4$

B：X 的眾數位置在 $X = 4$

C：機率 $P(X > 4) = e^{-1}$

D：平均數等於標準差

E：X 的分配是右偏的

Ans C, D, E

4 ：以下何種二項分配 $B(n, p)$，計算機率時，如使用常態近似二項機率，其結果可以被接受

A：$B(n = 20, p = 0.4)$

B：$B(n = 20, p = 0.8)$

C：$B(n = 20, p = 0.35)$

D：$B(n = 80, p = 0.95)$

E：$B(n = 120, p = 0.9)$

Ans A, C, E

5 ：已知班上 50 名同學中有 20 名女生 30 名男生。現在由班上隨機抽出一名同學，若該名學生為女性則令隨機變數 $X = 0$，若為男性則令 $X = 1$。則

A：X 之期望值 $E(X) = 0.4$

B：X 為 Bernoulli 分配

C：X 之變異數 $Var(X) = 0.6$

D：$P(X = 1) = E(X)$

E：$P(X = 0) = Var(X)$

Ans B, D

6 ：Austin 參加一個測驗，共有 25 題五選一單選題，若他全部用猜的，則

A：他猜中題數的期望值為 5

B：他猜中題數的變異數為 5

C：他猜中 5 題的機率最大。

D：他猜中 10 題的機率比猜中 3 題的機率大

E：他猜中 0 題的機率為 0.8^{25}

Ans A, C, E

7 ：某銀行在每一小時內顧客出現的人次為 $\mu = 6$ 的 Poisson 波氏分配。試問以 20 分鐘內來顧客人數為隨機變數 X，則

A：變異數 $Var(X) = 6$

B：期望值 $E(X) = 2$

C：$E(X) = Var(X)$

D：$E(X) + Var(X) = 2$

E：$E(X) \times Var(X) = 4$

Ans B, C, E

8 ：已知班上 50 名同學中有 20 名女生，30 名男生。現在由班上隨機抽出 10 名同學，且抽中男生人數為隨機變數 X、期望值為 $E(X)$ 與變異數為 $Var(X)$。則

A：$E(X) = 6$

B：$Var(X) = 2.4$

C：$E(X) = Var(X)$

D：$Var(X) < 2$

E：$E(X) - Var(X) < 2$

Ans A, D, E

9　：假設 X 的累積分配函數 $F(x)=\begin{cases}0, x<0\\ 1/4, 0\le x<1\\ x/3, 1\le x<2\\ 1, x\ge 2\end{cases}$，試問下列何者正確？

A：$P(X=1)=0$。

B：$P(X=2)=\frac{1}{3}$。

C：$P(1/2<X\le 2)=\frac{3}{4}$。

D：$P(1\le X<2)=\frac{1}{2}$。

E：$P(X=0)=\frac{1}{4}$。

Ans B, C

單　元

5

抽樣分配

在前面的章節中，我們討論了描述性統計及機率分配等相關主題，現在就要進入統計中最重要的一環－統計推論（Statistical Inference）。統計推論主要的概念是藉由樣本資料中所隱含的訊息來推論母體可能的特徵。一般多採用估計（Estimation）及假設檢定（Hypothesis Testing）兩種程序來進行統計推論；但因為統計推論相關理論與應用的層面太廣泛，我們將先介紹跟估計有關的專有名詞，然後再介紹抽樣後統計量的一些性質。

I. 重要名詞釋義

1. **估計量 (式)** (Estimator)：

   用來估計母體未知參數的統計量，稱之為估計量 (式)。

2. **估計值** (Estimate)：

   將樣本資料代入估計量之後所得到的值，稱之為估計值。

3. **點估計量** (Point Estimator)：

   由母體中隨機抽出樣本，而代入某一個統計量，來對母體參數進行估計，則此統計量稱為點估計量 (例如：$\overline{X}$、S^2)。

4. **點估計值** (Point Estimate)：

   由母體中隨機抽樣，而將樣本資料代入某一個統計量，計算所得到的值，稱之為點估計值 (例如：$\overline{X}$ = 70)。

5. **抽樣分配** (Sampling Distribution)：

   統計量 (Statistic) 的機率密度函數。

   估計方式又可細分成點估計與區間估計兩種形式來進行。

II. 點估計量的評估

如何知道一個點估計式的“好”與“壞”呢？回答這問題之前，那就要先定義什麼叫做“好”，下面是幾個常用來評估點估計式性質的準則。

1. 不偏性 (Unbiasedness)

下面是不偏性的定義。

定義 5.1 統計量$\hat{\theta}$是母體參數θ的估計式，若$\hat{\theta}$的期望值$E(\hat{\theta})=\theta$，則我們稱$\hat{\theta}$是θ的不偏估計式。

不偏估計式的例子不勝枚舉，例如：$\overline{X}$ 是 μ 的不偏估計式，S^2 是 σ^2 的不偏估計式，即$E(\overline{X})=\mu$及$E(S^2)=\sigma^2$，也就是說，樣本變異數$S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$的期望值是母體變異數 σ^2。現在大家可以了解當初在定義樣本變異數時為何要除以 $n-1$ 而不除以 n 了嗎？

不過採用不偏估計式的缺點，是它並不唯一，而且它不一定存在，會造成使用上的困擾。下面我們就要來介紹另一種準則。

2. 有效性 (Efficiency)

若$\hat{\theta}$是θ的估計式，則它的均方差 (Mean Square Error，*MSE*) 為

$$
\begin{aligned}
MSE(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 \\
&= \mathrm{E}(\hat{\theta}-\mathrm{E}(\hat{\theta})+\mathrm{E}(\hat{\theta})-\theta)^2 \\
&= \mathrm{E}(\hat{\theta}-\mathrm{E}(\hat{\theta})\ ^2+(\mathrm{E}(\hat{\theta})-\theta)^2 \\
&= Var(\hat{\theta})+(bias)^2
\end{aligned}
$$

均方差 *MSE* 是一種衡量誤差的準則，誤差當然是愈小愈好，所以 *MSE* 也可以用來評估估計式，下面是相對有效估計式的定義。

定義 2.2 對任意兩個θ的估計式$\hat{\theta}_1$及$\hat{\theta}_2$，若

$$\frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} < 1$$

則稱$\hat{\theta}_1$為相對有效估計式 (Relatively Efficient Estimator)。

另外，大家應該可以注意到不偏估計式的偏差 (Bias) 為 0，所以它的 MSE 就等於它的變異數，在選擇估計式時，較小的 MSE 在不偏估計式中，等價於較小的變異數。而在不偏性中，我們提到不偏估計式的不唯一性，所以在眾多的不偏估計式中以有效性來評估時，可以分成兩方面來探討。

定義 5.3 假設在所有的不偏估計式$\hat{\theta}$中，若$\hat{\theta}^*$具下列性質：

$$E(\hat{\theta}^*) = \theta \text{，且} Var(\hat{\theta}^*) \le Var(\hat{\theta}) \text{ , } \forall \text{不偏估計式} \hat{\theta} \text{，}$$

則$\hat{\theta}^*$稱為最有效率估計式 (The Most Efficient Estimator)。

定義 5.4 若$\hat{\theta}_1$和$\hat{\theta}_2$皆為θ的不偏估計式，即 $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ 且 $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} < 1$

則$\hat{\theta}_1$是θ的相對有效估計式。

3. 一致性 (Consistency)

定義 5.5 若一個估計式具備下列性質時：

當樣本數增加時，其所產生的估計式接近母體參數真值的機率也相對提高，則我們稱這樣的估計式為一致估計式 (Consistent Estimator)。

III. 抽樣分配

估計主要的依據是統計量的抽樣分配，因為本單元是以母體平均數為主，而母體平均數 μ 的估計多以樣本平均數 $\overline{X}$ 為出發點，所以我們只介紹有關樣本平均數 $\overline{X}$ 抽樣分配的定理。

定理 5.1 假設隨機變數 X 的平均數為 μ，標準差為 σ。若從中抽取一組大小為 n 的樣本，得其樣本平均數 $\overline{X}=\sum_{i=1}^{n} X_i/n$，則

(1) $\overline{X}$ 的平均數為 μ，標準差為 $\sigma/\sqrt{n}$。

(2) 若隨機變數 X 服從常態分配，則 $\overline{X}\sim N(\mu,\sigma/\sqrt{n})$。

(3) 若隨機變數 X 原始分配未知，則只要樣本數夠大 ($n\to\infty$)，則 $\overline{X}$ 仍會近似於 $N(\mu,\sigma/\sqrt{n})$。

定理 5.1 中的 (3)，其實就是有名的中央極限定理 (Central Limit Theorem)，其中樣本數的限制是 $n\to\infty$，應用上幾乎是不可行的；不過只要取 $n\geq 30$，$\overline{X}$ 的抽樣分配就不至於太偏離 $N(\mu,\sigma/\sqrt{n})$；所以在實務上只要是 $n\geq 30$，則會被視同是“大”樣本，而採用中央極限定理的結論來進行統計推論。所以透過 $\overline{X}$ 的抽樣分配，我們就可以計算其發生的機率，例如 $P(a<\overline{X}<b)$ 的機率可用下列方式求得：

$$P(a<\overline{X}<b)=P(\frac{a-\mu}{\sigma/\sqrt{n}}\leq\frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}}\leq\frac{b-\mu}{\sigma/\sqrt{n}})=P(\frac{a-\mu}{\sigma/\sqrt{n}}\leq Z\leq\frac{b-\mu}{\sigma/\sqrt{n}})$$

最後利用標準常態表求出標準常態隨機變數 Z 介於 $\frac{a-\mu}{\sigma/\sqrt{n}}$ 及 $\frac{b-\mu}{\sigma/\sqrt{n}}$ 間之機率即可。

範例 5.1 假設一般高中生每月的手機通訊費用的標準差為 \$360，平均費用為 \$1000，某月份隨機抽取 36 名高中生，試問此 36 名高中生當月手機通訊平均費用介於 \$1120 和 \$1180 的機率為何？

Ans

$$P(1120<\overline{X}<1300)=P(\frac{1120-1000}{360/\sqrt{36}}\leq\frac{(\overline{X}-1000)}{360/\sqrt{36}}\leq\frac{1180-1000}{360/\sqrt{36}})=P(2\leq Z\leq 3)$$

$$=P(2\leq Z\leq 3)=P(0<Z<3)-P(0<Z<2)=0.4987-0.4772=0.0215$$

演練題 (Q & A)

一、是非題

1 ：樣本平均數用來估計母體平均數，雖具有不偏性，但有可能不是最有效的估計。

Ans X

2 ：當樣本數不斷增加時，則樣本平均數的極端值將會逐漸減小。

Ans O

3 ：從成功率 p = 0.4 的貝努利 (Bernoulli) 母體獨立抽取 25 個樣本，則樣本成功比率的抽樣分配會近似於常態。

Ans O

4 ：從一個平均數為 10 的指數分配母體中，簡單隨機抽取 n = 40 個樣本，則抽出的樣本平均數的抽樣分配仍為近似指數分配。

Ans X

5 ：從一個母體個數 N = 100 中，以取後不放回的方式抽取了百分之十的樣本，並計算出樣本標準差，則有限母體校正係數為 0.9535。

Ans O

二、選擇題

1 ：從一個平均數 μ = 40 且變異數 σ^2 = 100 的連續母體中抽取 n = 100 個樣本的，則所有樣本平均數落入 38.04 至 41.96 的機率約為：

Standard Normal table: probability for P(0 < Z <z)										
z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：0.99

B：0.90

C：0.80

D：0.95

Ans D

2　：抽樣分配指的是以下何者的機率分配情形。

A：統計量 (Statistic)

B：母數 (Parameter)

C：母體 (Population)

D：統計量 (Statistic) 與母數 (Parameter) 兩者皆可

Ans A

3 ：從一個平均數 μ = 5 標準差為 σ = 4 的常態母體中，隨機抽取 16 個樣本，並計算樣本平均數。則樣本平均數在以下那一個區間範圍的機率約為 95%?

A：[4, 6]

B：[3, 7]

C：[2, 8]

D：[1, 9]

Ans B

4 ：中央極限定理說明當樣本數逐漸增加時，樣本平均數會越來越趨向於 _______ 分配。

A：卡方

B：T

C：均等

D：常態

Ans D

5 ：若某次抽樣樣本數為 40，得到樣本平均數的標準誤為 20，現在希望在相同取樣情形下，將標準誤減少到 10，需要抽取的樣本數應該為何？

A：將樣本數增加到 60

B：將樣本數增加到 80

C：將樣本數增加到 160

D：將樣本數減少到 20

Ans C

6　：樣本平均數的期望值等於母體平均數，這是以下哪一種的性質。

A：不偏性 (Unbiased)

B：最小變異性 (Minimum Variance)

C：有效性 (Efficiency)

D：一致性 (Consistency)

Ans A

7　：假設參加統計學線上測驗的考生，其期望年齡及標準差分別是 24 歲與 4 歲。現隨機抽取 64 名考生，則此樣本平均數落在 [23, 25] 歲區間的機率為何？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

A：0.4772

B：0.9544

C：0.9015

D：0.9605

Ans B

8 ：假設參加統計學線上測驗的考生，其期望年齡及標準差分別是 24 歲與 4 歲。現隨機抽取 64 名考生，則此樣本平均數是何種分配。

A：近似常態分配

B：T 分配

C：卡方分配

D：指數分配

Ans A

9 ：某螺絲釘工廠所生產的螺絲平均長度為 1 公分，標準差為 0.02 公分。隨機抽取 36 支螺絲釘，則樣本平均長度的標準誤為多少公分？

A：0.0033

B：0.02

C：0.002

D：0.005

Ans A

10 ：某螺絲釘工廠所生產的螺絲平均長度為 1 公分，標準差為 0.02 公分。隨機抽取 36 支螺絲釘，則樣本平均長度的超過 1.01 公分的機率為何？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：0.0013

B：0.025

C：0.3085

D：0.052

Ans A

11 ：當從有限母體中以取後不放回的方式隨機抽取樣本，則有限母體校正會使得樣本平均數之 ________?

A：標準誤會加大

B：標準誤沒有影響

C：標準誤會減小

D：標準誤有時增加有時減小

Ans C

12 ：當抽取了 n = 20 個樣本，得到樣本平均數的抽樣分配近似於常態分配，以下哪一種連續母體較為可能

A：任何母體均可

B：母體大致為對稱

C：母體為右偏分配

D：母體為常態分配

Ans B

13 ：估計母體平均數時，使用樣本平均數會比其他任何的中央趨勢估計其標準差都較小，此為何種性質？

A：不偏性 (Unbiased)

B：有效性 (Efficiency)

C：一致性 (Consistency)

D：變化性 (Variability)

Ans B

14 ：隨機投擲兩粒正常的骰子，並計算此隨機實驗的平均點數，則平均點數的期望值為何？

A：3

B：3.5

C：4

D：7

Ans B

15 ：隨機投擲兩粒正常的骰子，並計算此隨機實驗的平均點數，則平均點數的變異數為何？

A：3.5

B：4.2

C：35/24

D：35/12

Ans C

16 ：矽石公司生產的電池平均壽命為 300 小時，標準差為 100 小時，今某人隨機購買 64 個此種電池，試問這 64 個電池平均壽命的抽樣分配大約為何種分配？

A：常態分配

B：T 分配

C：卡方分配

D：無法確定

Ans A

17 ：矽石公司生產的電池平均壽命為 300 小時，標準差為 100 小時，今某人隨機購買 64 個此種電池，試問這 64 個電池平均壽命的標準差為何？

A：100 小時

B：64 小時

C：12.5 小時

D：1.5625 小時

Ans C

18 ：中央極限定理說明當樣本數充分大時，何種統計量的抽樣分配會愈來愈趨於常態？

A：樣本平均數

B：樣本標準差

C：樣本中位數

D：樣本最大值

Ans A

19 ：有個不公正的銅板，正反面出現的機率分別為 0.7 及 0.3，現在投擲此銅板 42 次，得到銅板正面平均數的期望值為多少？

A：0.7

B：0.3

C：0.5

D：3.5

Ans A

20 ：有個不公正的銅板，正反面出現的機率分別為 0.7、0.3，現在投擲此銅板 42 次，得到銅板正面平均數的變異數為多少？

A：0.21

B：8.82

C：0.021

D：0.005

Ans D

21 ：從一個在 10 至 16 間連續均等分配 U(10, 16) 中，隨機抽取 36 個樣本，形成樣本平均數。則此樣本平均數的期望值為何？

A：13

B：10

C：16

D：18

Ans A

22 ：從一個在 10 至 16 間連續均等分配 U(10, 16) 中，隨機抽取 36 個樣本，以此形成樣本平均數。則此樣本平均數的變異數為何？

A：3

B：9

C：1/12

D：13/36

Ans C

23 ：從一個在 10 至 16 間連續均等分配 U(10, 16) 中，隨機抽取 36 個樣本，以此形成樣本平均數。則此樣本平均數大約是何種分配。

A：無法決定

B：卡方分配

C：均等分配

D：常態分配

Ans D

24 ：從一個在 10 至 16 間連續均等分配 U(10, 16) 中，隨機抽取 36 個樣本，以此形成樣本平均數。則此樣本平均數大於 14 的機率為何？

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：0.0002

B：0.1587

C：0.3413

D：0.025

Ans A

三、複選題

1　：利用樣本平均數來估計母體平均數時，其具有那些特性？

A：不偏性

B：有效性

C：一致性

D：不變性

E：最大性

Ans A, B, C

2　：從一個平均數為 10 的指數分配母體中，隨機抽取 n = 100 個樣本，則抽出的樣本平均數有哪些特性？

A：樣本平均數抽樣分配近似常態分配

B：樣本平均數的期望值為 10

C：樣本平均數的標準差為 10

D：樣本平均數落在 [8, 12] 的機率約為 95%

E：樣本平均數大於 20 的機率約為 e^{-2}

Ans A, B, D

3　：隨機投擲兩粒正常的骰子，並計算此隨機實驗的平均點數，則

A：平均點數的期望值為 3.5

B：平均點數的變異數為 35/12

C：平均點數大於 5 的機率為 1/12

D：平均點數的抽樣分配仍為均等分配

E：平均點數等於 3.5 的機率為 1/6

Ans A, C, E

4 ：假設參加統計學線上測驗的考生，其期望年齡及標準差分別是 24 歲與 4 歲。現隨機抽取 64 名考生，則

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：樣本平均數落在 [23, 25] 歲區間的機率約為 0.95

B：樣本平均數落在 [22, 26] 歲區間的機率約為 0.3830

C：樣本平均數標準誤為 0.5 歲

D：樣本平均數的期望值為 24 歲

E：樣本平均數的抽樣分配為 T 分配

Ans A, C, D

5　：從一個在 10 至 16 間連續均等分配 U(10, 16) 中，隨機抽取 36 個樣本，並形成樣本平均數。則

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

A：此樣本平均數是近似常態分配

B：樣本平均數的標準誤為 3

C：樣本平均數的期望值為 13

D：樣本平均數落在 [12, 14] 區間的機率約為 1

E：樣本平均數小於 12 的機率約為 0.1587

Ans A, C, D

單　元

6

信賴區間估計

在前面的章節中，我們討論了點統計量的性質及其抽樣分配等相關問題，而在這個單元中，我們將介紹估計的另一種形式：區間估計 (Interval Estimation) 。由於涵蓋面非常廣，我們將先討論單一及兩母體平均數區間估計的範疇，然後再介紹單一及兩母體比例的區間估計問題。

I. 重要名詞釋義

1. 區間估計量 (Interval Estimator)：

由二個統計量構成一個區間，用此區間來估計母體參數可能落入的範圍，則此區間統計量稱之為區間估計量。

2. 區間估計值 (Interval Estimate)：

由母體中隨機抽樣，而將樣本資料代入區間統計量後，計算所得到的值，即為區間估計值。

3. 信賴區間 (Confidence Interval)：

設母體參數為θ，若從母體中抽樣，得估計θ的兩個統計量$\hat{\theta}_1$與$\hat{\theta}_2$，使得$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1-\alpha$，則稱$(\hat{\theta}_1, \hat{\theta}_2)$為$\theta$之 100(1−$\alpha$)% 信賴區間。

4. 信賴係數 (Confidence Coefficient)：

1−α 稱之為信賴係數，是用來衡量信賴區間之準確性。

II. 單一及兩母體平均數之區間估計

區間估計主要之依據是統計量的抽樣分配，根據定理 5.1(3) 中的中央極限定理，我們可以透過$\overline{X}$的抽樣分配（如圖 6.1 所示意的），建構出母體平均數 μ 的 100(1−α)% 區間估計式：

$$P(-Z_{\alpha/2} \le \frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}} \le Z_{\alpha/2}) = 1-\alpha$$

$$\Rightarrow P(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

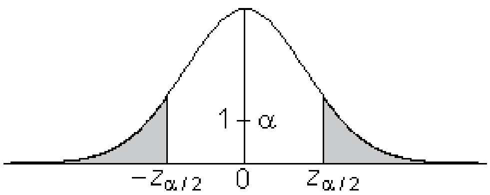


圖 6.1 Z 分配機率示意圖

另外，樣本數不夠大 (實務上 n < 30) 時，則應分成兩方面來進行探討，若已知 X 是服從常態分配 $N(\mu, \sigma)$，其中平均數 μ 未知而標準差 σ 已知時，從定理 5.1 可以得知母體平均數 μ 的建構方式同以上所介紹大樣本的方法一樣。但若已知 X 是服從常態分配，其中平均數 μ 和標準差 σ 皆未知，一般作法是採用樣本標準差 S 來估計標準差 σ，則此時 $\frac{(\overline{X}-\mu)}{S/\sqrt{n}}$ 就不再服從常態分配，而是服從自由度為 $n-1$ 的 t 分配了。至於在小樣本非常態分配的狀況下，本單元所介紹的方法並不適用，應採用其他方法（例如無母數方法等）來回答。下面我們將依照不同母體條件下，針對母體平均數區間估計分別加以介紹。

II-1. 單一母體平均數之區間估計

單一母體平均數之估計彙整如表 6.1。

表 6.1 單一母體平均數 μ 之區間估計

參數	條件		100(1−α)% 區間估計式
	樣本大小	母體變異數	
μ	大	已 知	$\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
		未 知	若母體變異數未知，以樣本變異數估計之
	小 (Normal)	已 知	
		未 知	$\overline{X} \pm t_{\alpha/2,(n-1)}\frac{S}{\sqrt{n}}$

II-1-1. 大樣本

1. 單一母體平均數 + 大樣本 + σ^2 已知

範例 6.1 假設一般高中生每月的手機通訊費用的標準差為 \$300，某月份隨機抽取 36 名高中生得當月手機通訊平均費用為 \$1100：

(1) 請決定高中生每月手機通訊平均費用的 90% 信賴區間。

(2) 若想要達到 95% 的信賴水準，最大誤差不超過 \$150，所需要蒐集的樣本數為何？

816	691	1209	1061	1052	683	1054	842	1096
1623	1213	979	1210	960	1386	1421	978	1192
936	695	753	1239	1233	1021	1846	1282	1209
1234	597	1364	1204	937	1345	1389	854	996

Ans

(1) 令 $\overline{X}$ 為高中生某月的手機通訊平均費用，則

$$\bar{X} \pm Z_{0.05}\frac{\sigma}{\sqrt{n}} \Rightarrow 1100 \pm 1.645 \times \frac{300}{\sqrt{36}} = 1100 \pm 1.645 \times 50 = 1100 \pm 82.25$$

∴高中生每月的手機通訊平均費用之 90% 信賴區間為 (1017.75, 1182.25)。

(2) $n=(\frac{Z_{\alpha/2}\sigma}{E})^2 = (\frac{Z_{0.025}\sigma}{150})^2 = (\frac{1.96\times 300}{150})^2 = 15.3664 \rightarrow 16$ 位高中生。

2. 單一母體平均數 + 大樣本 + σ^2 未知

範例 6.2 某醫療研究機構想要估計平均罹患心臟病的年齡，因此從心臟病患中隨機抽取 36 人得下列資料 (歲)：

54	09	41	47	57	75	68	59	72	77	61	38	45	81	25	49	60	12
66	37	58	62	49	54	62	64	03	74	51	71	79	27	69	60	83	81

其中 $\bar{x}=55$，$s=20.49$。

求罹患心臟病平均年齡的 95% 信賴區間。

Ans

令$\overline{X}$為平均罹患心臟病的年齡，則

$$\overline{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}} \Rightarrow 55 \pm 1.96\frac{20.49}{\sqrt{36}} = 55 \pm 6.6934 \Rightarrow (48.3066,61.6934)$$

∴平均罹患心臟病年齡之 95% 信賴區間為 (48.3066, 61.6934)。

II-1-2. 小樣本

1. 單一母體平均數 + 小樣本 +Normal + σ^2 已知或未知

範例 6.3 某農場出產的牛奶，每瓶重量呈標準差為 6 公克的常態分配。消基會抽檢 16 瓶該農場的牛奶，得下列資料及其平均重量為 510 公克，試問：

511	508	513	510	516	502	508	507
508	513	518	513	510	515	508	500

(1) 請問該農場的牛奶每瓶的平均重量之 95% 信賴區間為何？

(2) 若母體標準差未知 (而樣本標準差已知為 4.775 公克)，請重複回答 (1)。

Ans

(1) 令$\overline{X}$為牛奶每瓶的平均重量，則

$$\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Rightarrow 510 \pm 1.96\frac{6}{\sqrt{16}} = 510 \pm 2.94$$

∴該農場的牛奶每瓶的平均重量之 95% 信賴區間為 (507.06,512.94)

(2) 令$\overline{X}$為牛奶每瓶的平均重量，則

$$\overline{X} \pm t_{\alpha/2,(n-1)} \frac{S}{\sqrt{n}} \Rightarrow 510 \pm 2.13 \frac{4.775}{\sqrt{16}} = 510 \pm 2.5427$$

∴該農場的牛奶每瓶的平均重量之 95% 信賴區間為 (507.4573,512.5427)

II-2. 兩母體平均數差之區間估計

假設隨機變數 X_1、X_2 分別來自平均數為 μ_1、μ_2，標準差為 σ_1、σ_2 之母體。若各從中抽取一組大小為 n_1、n_2 的樣本，得其樣本平均數 $\overline{X}_1$ 、 $\overline{X}_2$ ，樣本變異數 S_1^2、S_2^2 則兩母體平均數差 $\mu_1-\mu_2$ 之區間估計彙整如表 6.2。

表 6.2 兩母體平均數差 $\mu_1-\mu_2$ 之區間估計

<table>
<tr><th rowspan="2">參數</th><th colspan="3">條件</th><th rowspan="2">100(1−α)% 區間估計式</th></tr>
<tr><th>樣本大小</th><th colspan="2">母體變異數</th></tr>
<tr><td rowspan="5">$\mu_1-\mu_2$</td><td rowspan="2">大</td><td colspan="2">已 知</td><td rowspan="3">$(\overline{X}_1-\overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
若母體變異數未知，以樣本變異數估計之</td></tr>
<tr><td colspan="2">未 知</td></tr>
<tr><td rowspan="3">小
(Normal)</td><td colspan="2">已 知</td></tr>
<tr><td rowspan="2">未
知</td><td>相
等</td><td>$(\overline{X}_1-\overline{X}_2) \pm t_{\frac{\alpha}{2},(n_1+n_2-2)}\sqrt{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})}$ ，
其中 $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$</td></tr>
<tr><td>不
等</td><td>$(\overline{X}_1-\overline{X}_2) \pm t_{\frac{\alpha}{2},(v)}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$
其中 $v = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1}+\frac{\left(S_2^2/n_2\right)^2}{n_2-1}}$</td></tr>
</table>

參數	條件		100(1−α)% 區間估計式
	樣本大小	母體變異數	
μ_D	小 (Normal)	未 知	$\bar{d} \pm t_{\frac{\alpha}{2},(n-1)} \frac{S_d}{\sqrt{n}}$，其中，$d_i = X_{1i} - X_{2i}$， $\bar{d} = \frac{\sum d_i}{n}, S_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2 / n}{n-1}}$

II-2-1. 兩大樣本之區間估計

1. 兩母體平均數差 + 大樣本 + σ_1^2, σ_2^2 已知或未知

範例 6.4 根據某人力資源雜誌隨機抽樣的結果得知今年大學畢業生平均薪資，在 34 位男大學畢業生中，其起薪資料為 $\bar{x}_{男} = \$28{,}121$，$s^2_{男} = 28476230$；而 30 位女大學畢業生中，其起薪資料為 $\bar{x}_{女} = \$23{,}643$，$s^2_{女} = 25213057$。請估計男大學畢業生平均起薪和女大學畢業生平均起薪差的 98%信賴區間。

Ans 母體 1: 男大學畢業生薪資 $n_{男} = 34$, $\bar{x}_{男} = 28121$, $s^2_{男} = 28476230$

母體 2: 女大學畢業生薪資 $n_{女} = 30$, $\bar{x}_{女} = 23643$, $s^2_{女} = 25213057$

$$\bar{X}_{男} - \bar{X}_{女} \pm Z_{\alpha/2}\sqrt{\frac{S^2_{男}}{n_{男}} + \frac{S^2_{女}}{n_{女}}}$$

$$\Rightarrow (28121 - 23643) \pm Z_{0.01}\sqrt{\frac{28476230}{34} + \frac{25213057}{30}}$$

$$\Rightarrow 4478 \pm 2.33 \times (1295.365) \Rightarrow \text{約 } 4478 \pm 3018.2$$

所以男大學畢業生平均起薪和女大學畢業生平均起薪差的 98% 信賴區間為 (\$1459.8, \$7496.2)。

II-2-2. 兩小樣本之區間估計

1. 兩母體平均數差 + 小樣本 + Normal + σ_1^2, σ_2^2 已知

範例 6.5 想要比較澳洲小朋友與台灣小朋友每天運動時間是否有差異，分別隨機抽取 25 個澳洲小朋友，得 $\bar{x}_1 = 66.8$ 分，(假設已知澳洲小朋友每天運動的時間呈常態分配，標準差 $\sigma_1 = 16.88$ 分)，另外，亦隨機抽取台灣小朋友 35 位，得 $\bar{x}_2 = 46.29$ 分，(假設已知台灣小朋友每天運動的時間亦呈常態分配，標準差 $\sigma_2 = 14.26$ 分)，請估計兩國家小朋友平均每天運動時間差的 99%信賴區間。

Ans 母體 1: 澳洲小朋友每天運動時間 $n_1 = 25,\ \bar{x}_1 = 66.8,\ \sigma_1 = 16.88$

母體 2: 台灣小朋友每天運動時間 $n_2 = 35,\ \bar{x}_2 = 46.29,\ \sigma_2 = 14.26$

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \Rightarrow (66.8-46.29) \pm Z_{0.005}\sqrt{\frac{16.88^2}{25}+\frac{14.26^2}{35}}$$

$$\Rightarrow 20.51 \pm 2.58 \times 4.1482 \Rightarrow 20.51 \pm 10.7024$$

所以澳洲小朋友和台灣小朋友每天平均運動時間差的 99% 信賴區間為 (9.8076,31.2124)。

2. 兩母體平均數差 + 小樣本 + Normal + σ_1^2, σ_2^2 未知且不相等

範例 6.6 下列樣本資料為兩個廠區每月之用電度數

廠區 1	101	100	98	78	75	90	88	95	90	92
廠區 2	90	88	90	75	70	80	80	82	78	80

若已知兩個廠區用電量皆近似於常態分配，請估計廠區 1 的平均每月用電量與廠區 2 的平均月用電量差的 99%信賴區間。

Ans 廠區 1: $n_1 = 10,\ \bar{x}_1 = 90.7, s_1 = 8.71$

廠區 2: $n_2 = 10,\ \bar{x}_2 = 81.3, s_2 = 6.50$

令 μ_1 為廠區 1 的平均月用電量，μ_2 為廠區 2 的平均月用電量

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2,v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} = (90.7-81.3) \pm t_{0.005,17}\sqrt{\frac{8.71^2}{10}+\frac{6.50^2}{10}}$$

$$v = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\left(\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}\right)} = \frac{(8.71^2/10+6.50^2/10)^2}{\left(\frac{(8.71^2/10)^2}{10-1}+\frac{(6.50^2/10)^2}{10-1}\right)} \approx \frac{139.5094}{8.378} \approx 16.651 \to 17$$

$$\Rightarrow\ 9.4 \pm 2.898 \times 3.4368 \Rightarrow\ 9.4 \pm 9.9598$$

所以廠區 1 與廠區 2 的平均月用電量差的 99% 信賴區間為 (-0.5598, 19.3598)。

3. 兩母體平均數差 + 小樣本 + Normal + σ_1^2, σ_2^2 未知但相等

範例 6.7 想要了解在家自學與傳統教學法的學習成效是否有差異，教育部從在家自學的學生中，隨機抽取 25 名學生，得其平均學業表現分數為 76.96 分，標準差為 12.27 分；而傳統教學法的 60 名學生樣本中，其平均學業表現分數為 56.41 分，標準差為 15.58 分，另假設兩抽樣母體資料為常態分配且變異數相等。請估計在家自學學生的平均學業分數與傳統教學法學生的平均學業分數差的 99%信賴區間。

Ans 母體 1: 在家自學法學生的學業表現分數 $n_1 = 25$，$\bar{x}_1 = 76.96$，$s_1 = 12.27$

母體 2: 傳統教學法學生的學業表現分數 $n_2 = 60$，$\bar{x}_2 = 56.41$，$s_2 = 15.58$

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(25-1)\times 12.27^2+(60-1)\times 15.58^2}{25+60-2} \approx 216.08$$

$$\left(\bar{X}_1-\bar{X}_2\right) \pm t_{\alpha/2,(n_1+n_2-2)}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\Rightarrow \quad (76.96-56.41) \pm t_{0.005,83}\sqrt{216.08\left(\frac{1}{25}+\frac{1}{60}\right)}$$

$$\Rightarrow \quad 20.55 \pm 2.6364 \times 3.5 \quad \Rightarrow \quad 20.55 \pm 9.2274$$

所以在家自學學生的平均學業分數與傳統教學法學生的平均學業分數差的 99%信賴區間為 (11.3226, 29.7774)。

II-2-3. 成對樣本之區間估計

成對樣本之區間估計可以透過資料的處理而轉化成單一母體平均數的估計問題。假設從隨機變數 X_1、X_2 中抽取一組大小為 n 的成對樣本，令 $d_i = X_{1i} - X_{2i}$，$i = 1, 2, \cdots, n$，得其樣本平均數 $\bar{d} = \frac{\sum d_i}{n}$，樣本標準差 $S_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1}}$，則兩母體成對樣本平均數差 $\mu_D = \mu_1 - \mu_2$ 之區間估計公式如表 6.2 最後一個欄位所示。

1. 小樣本 + Normal + σ 未知 + 成對樣本

範例 6.8 某學者想要測試新藥是否具減肥效果。該實驗室使用 10 隻小白鼠當成樣本進行測試得下列資料：

	1	2	3	4	5	6	7	8	9	10
食用藥品前體重 (1)	5.0	4.7	6.6	7.0	6.7	4.7	5.7	6.0	7.4	5.1
食用藥品後體重 (2)	4.8	4.9	6.2	6.9	6.8	4.4	5.7	5.6	6.9	4.7

假設成對樣本平均數差構成之母體為常態，請估計食用藥品前平均體重與食用藥品後平均體重差的 95%信賴區間。

Ans

	1	2	3	4	5	6	7	8	9	10
食用藥品前體重 (1)	5.0	4.7	6.6	7.0	6.7	4.7	5.7	6.0	7.4	5.1
食用藥品後體重 (2)	4.8	4.9	6.2	6.9	6.8	4.4	5.7	5.6	6.9	4.7
(1)−(2) = d_i	0.2	-0.2	0.4	0.1	-0.1	0.3	0	0.4	0.5	0.4

$$\sum d_i = 2\,,\ \sum d_i^2 = 0.92\,,\ \bar{d} = \frac{2}{10} = 0.2\,,\ S_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1}} = 0.24$$

$$\bar{d} \pm t_{\alpha/2,\,(n-1)} \frac{S_d}{\sqrt{n}} = 0.2 \pm t_{0.025,9} \times \frac{0.24}{\sqrt{10}} = 0.2 \pm 2.262 \times 0.0759 = 0.2 \pm 0.1717$$

所以食用藥品前平均體重與食用藥品後平均體重差的 95％信賴區間為 (0.0283, 0.3717)。

III. 單一及兩母體比例之區間估計

在民意調查或市場調查等相關的議題上，我們常會遇到母體比例相關的統計推論問題。例如說，推論某位立委候選人的支持率或者研究某品牌的市佔率等，而母體比例統計推論的議題，我們也是將它分成單一母體與兩母體來探討；另外亦會介紹如何決定抽樣樣本數的大小。

在比例相關議題上，每一筆觀測值都只有“是”或“否”這兩種可能，以支持率為例來說，每一個受訪者被問到“你是否支持該位立委候選人？”時，他（她）的答案 (Y_i)：只有“支持”($Y_i = 0$) 或“不支持”($Y_i = 0$)，所以每個人的答案 (Y_i) 是一個隨機變數且服從 Bernoulli(p) 分配，而 p 是母體中真正支持該位立委候選人的比例；若我們訪問了 n 位得到 n 筆觀測值，就如同是執行了 n 次 Bernoulli(p) 的試驗，那麼 $X = \sum_{i=1}^{n} Y_i$ 就代表成功的次數且會服從 $B(n, p)$ 分配。

另外，我們也會用樣本比例 $\hat{p} = X / n$ 來估計母體的比例，如圖 6.2 所示，在進行母體比例的統計推論時，多會藉助 $\hat{p}$ 來建構信賴區間或進行假設檢定。

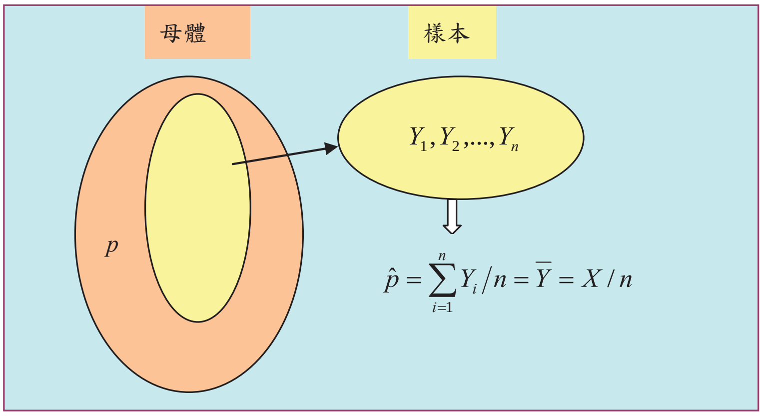


圖 6.2 母體比例的估計示意圖

III-1. 單一母體比例的區間估計

若 Y_i 是服從 Bernoulli(p)，則 $E(Y_i) = p$，$Var(Y_i) = p(1-p)$，因為 $\hat{p} = \sum_{i=1}^{n} Y_i / n = \overline{Y}$，所以根據中央極限定理可以得知，當樣本數很大時，

$$\hat{p} \overset{\text{近似}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

而這裡的所謂大樣本實務上可以是 $n \geq 30$，或是 $n \cdot p \geq 5$ 和 $n(1-p) \geq 5$。另外 $\hat{p}$ 的期望值 $E(\hat{p}) = p$，代表 $\hat{p}$ 是 p 的不偏估計式，因此 $\hat{p}$ 也是 p 最常用的點估計式。

所以透過 $\hat{p}$ 的抽樣分配，我們可以建構母體比例 p 的 $100(1-\alpha)\%$ 的信賴區間如下，

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

範例 6.9 某民調公司最近做了一項縣市長支持度調查，結果顯示在 400 名受訪者中有 200 名支持 1 號候選人，在 90% 的信賴水準下請估計，全部選民支持 1 號候選人的比例在何範圍之內？

Ans p：支持 1 號候選人的比例

$$n = 400, \quad x = 200 \quad , \quad \hat{p} = \frac{X}{n} = \frac{200}{400} = 0.5$$

而 p 的 90% 信賴區間為

$$\begin{aligned}
&\hat{p} \ \pm \ Z_{\alpha/2} \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \\
&\Rightarrow 0.5 \ \pm \ 1.645 \sqrt{\frac{0.5\left(1-0.5\right)}{400}} \\
&\Rightarrow 0.5 \pm 0.041 \\
&\Rightarrow \left(0.459\ , 0.541\right)
\end{aligned}$$

所以在 90% 的信賴水準下，該縣市選民支持 1 號候選人的比例落在 (0.459, 0.541) 之間。

III-2. 兩母體比例差之區間估計

兩個母體比例統計推論的相關應用，比如說比較兩個政黨對某一項政策的支持率，或者比較兩種品牌的市佔率，或者比較兩條生產線所製造的產品良率等，都是屬於兩個母體比例的範疇。

在進行兩母體比例的估計與假設檢定時，需分別從母體 1 及母體 2 中蒐集兩獨立樣本的資料，其中樣本比例 $\hat{p}_1$ 及 $\hat{p}_2$ 是推論的重要依據，尤其我們多是透過 $\hat{p}_1 - \hat{p}_2$ 的抽樣分配來建構兩母體比例差之信賴區間及進行假設檢定。

若從母體 1 中抽出 (大樣本) n_1 個樣本點，其中成功的次數為 X_1，且 $\hat{p}_1 = \sum_{i=1}^{n_1} Y_i / n_1 = X_1 / n_1$ 是母體比例 p_1 的估計式，則同前面單一母體比例部分所解釋

的，$\hat{p}_1$ 會近似於 $N(p_1, p_1(1-p_1)/n_1)$。同樣的，若從母體 2 中抽出 (大樣本) n_2 個樣本點，其中成功的次數為 X_2，且 $\hat{p}_2=\sum_{i=1}^{n_2} Y_i/n_2 = X_2/n_2$ 是母體比例 p_2 的估計式，則 $\hat{p}_2$ 會近似於 $N(p_2, p_2(1-p_2)/n_2)$。

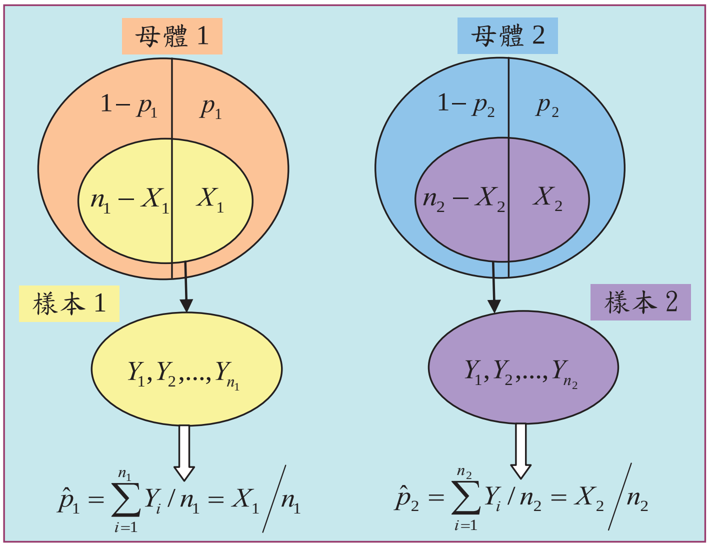


圖 6.3 兩母體比例估計示意圖

那麼 $\hat{p}_1-\hat{p}_2$ 的抽樣分配為

$$\hat{p}_1-\hat{p}_2 \overset{\text{近似}}{\sim} N\left(p_1-p_2, \frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}\right)$$

因此如圖 6.3 所示意的，可以得知：

$$P\left(-Z_{\alpha/2} < Z = \frac{\hat{p}_1-\hat{p}_2-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} < Z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\hat{p}_1-\hat{p}_2-Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}} < p_1-p_2\right.$$

$$< \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\Bigg) = 1-\alpha$$

所以兩母體比例差 $\hat{p}_1 - \hat{p}_2$ 的 $100(1-\alpha)\%$ 信賴區間為

$$\left(\hat{p}_1 - \hat{p}_2 - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}},\ \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

範例 6.10 為明瞭某兩個地區對修建隧道的意向，特地由地區 1 和地區 2 分別隨機抽出居民 1000 人及 600 人加以調查，其中地區 1 和地區 2 回答贊成者分別為 750 人及 360 人，試建構兩地區居民贊成修隧道的比例差的 95% 信賴區間。

Ans p_1：代表地區 1 贊成修建隧道的比例

p_2：代表地區 2 贊成修建隧道的比例

地區 1：$n_1 = 1000$，$x_1 = 750$，$\hat{p}_1 = \dfrac{750}{1000} = 0.75$

地區 2：$n_2 = 600$，$x_2 = 360$，$\hat{p}_2 = \dfrac{360}{600} = 0.6$

$p_1 - p_2$ 的 95% 信賴區間為

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\Rightarrow (0.75-0.6) \pm 1.96\sqrt{\frac{(0.75)(0.25)}{1000} + \frac{(0.6)(0.4)}{600}}$$
$$\Rightarrow 0.15 \pm 0.048 \Rightarrow (0.102\ ,\ 0.198)$$

所以兩地區贊成修建隧道的比例差之 95% 信賴區間為 (0.102, 0.198)。

IV. 樣本數的選取

在進行調查時，若能取得愈多的樣本數，則資訊愈充足，誤差自然就會愈小。但一般的民調或市調經費大多有限，所以樣本數的大小要如何兼顧精確度及成本的要求，是一個不容忽視的問題。

若要求信賴係數為 $1-\alpha$ 且所能容忍的最大誤差界限為 E 時，在比例估計時

$$E = Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$\Rightarrow E^2 = Z_{\alpha/2}^2 \cdot \frac{p(1-p)}{n}$$

$$\Rightarrow n = p(1-p)\cdot\left(\frac{Z_{\alpha/2}}{E}\right)^2 \tag{1}$$

現在讓我們從有無事前資訊兩方面加以探討。

IV-1. 有事前資訊

因為 α 和可容忍最大誤差界限 E 的條件是事先給定的，若要透過公式 (1) 來決定樣本數 n 的大小，則額外需要 p 的資訊。一般來說母體比例 p 都是未知的，所以 p 的估計式 $\hat{p}$ 可以透過專家的意見直接給定，或者可以透過初步樣本所估計的樣本比例 $\hat{p}$ 來估計 p。也就是說

$$n = \hat{p}(1-\hat{p})\cdot\left(\frac{Z_{\alpha/2}}{E}\right)^2$$

IV-2. 無事前資訊

若無事前資訊時，為了要達到精確度的要求，即信賴係數為 $1-\alpha$ 且最大容忍誤差界限為 E 時，通常會在公式中取等號左邊 $p\cdot(1-p)$ 的最大值代入，至於 $p\cdot(1-p)$ 的最大值為何呢？答案是 1/4。理由是透過二階導數判別法，可以求得 $f(p) = p(1-p)$

的最大值是發生在 $p = 1/2$ 時，所以樣本數由下面公式決定

$$n = \frac{1}{4} \cdot \left(\frac{Z_{\alpha/2}}{E} \right)^2$$

註

求 $f(p) = p(1-p)$ 的極大值，做法如下：

$f'(p) = 1 - 2p$

令 $f'(p) = 0$ 臨界數 $p = 1/2$

$f''(p) = -2\big|_{p=1/2} < 0 \Rightarrow$ $f(p)$ 下凹，所以在 $p = 1/2$ 時，$f(p)$ 有極大值。

範例 6.11 行政院想要了解民眾對核四案的立場進行調查，若要求 95% 的信心水準且誤差不超過三個百分點，則

(1) 樣本數應取多少？

(2) 若事先預知，贊成的民眾約佔四成，則樣本數應取多少？

Ans 令 p：贊成興建核四廠的比例

(1) 無事前資訊

$$\therefore n = \frac{1}{4}\left[\frac{Z_{\alpha/2}}{E}\right]^2 = \frac{1}{4}\left[\frac{1.96}{0.03}\right]^2 \approx 1067.11 \to 1068$$

若無事先資訊，在要求 95% 的信心水準且誤差不超過 3 個百分點下，則樣本數至少應取 1068 份。

(2) 事前資訊 $\hat{p}_1 = 0.4$

$$n = \hat{p}(1-\hat{p})\left(\frac{Z_{\alpha/2}}{E}\right)^2 = 0.4(1-0.4)\left(\frac{1.96}{0.03}\right)^2 \approx 1024.42 \to 1025$$

$\therefore$若事先預知贊成的民眾約佔四成，在要求 95% 的信心水準且誤差不超過 3 個百分點下，則樣本數至少應取 1025 份。

演練題 (Q & A)

一、是非題

1 :在建構非獨立成對樣本資料平均差的信賴區間時，此信賴區間的中心點為 μ_d。

Ans X

2 ：假設常態母體平均值 μ 的信賴區間為 $\bar{x} \pm 1.645\sigma/\sqrt{n}$，則其信賴水準 (Confidence Level) 為 90%。

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans O

3 ：母體平均值 μ 的信賴區間，一定包含未知參數 μ 的真實值。

Ans X

4 ：通常母體平均值 μ 的信賴區間，其區間的中心點為 $\bar{x}$。

Ans O

5 ：兩母體比例差之信賴區間永遠介於 -1 和 1 之間。

Ans O

6 ：兩母體比例差之點估計為 $\hat{p}_1 - \hat{p}_2$，當分別抽樣之樣本數充分大時，其抽樣分配為常態分配。

Ans O

7 ：在估計母體比率時，需考慮隨機抽樣樣本數時，是利用 $n=(Z_{\alpha/2})^2 pq/E^2$ 來求算，若 p 的合理估計值無法獲得，則公式中的 p 可用 0.5 來代替。

Ans O

8 ：從兩母體隨機抽取樣本，則不管樣本數大小為何，$\bar{X}_1-\bar{X}_2$ 之抽樣分配為常態分配。

Ans X

9 ：從兩常態母體隨機抽取樣本，則不管樣本數大小為何，$\bar{X}_1-\bar{X}_2$ 之抽樣分配為常態分配。

Ans O

10 ：隨機抽取 49 個樣本，得其樣本平均值及標準差分別為 30 和 14，則母體平均值 95% 信賴區間的上限值約為 33.29。

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans X

11 ：隨機從常態母體中抽取 16 個樣本，得其樣本平均值及標準差為 8 和 9，則母體平均值 90% 信賴區間的下限值約為 6.77。

Critical value of t distribution					
degree of freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
27	1.314	1.703	2.052	2.473	2.771
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
Normal	1.282	1.645	1.960	2.326	2.576

Ans X

12 ：從有限母體 $N = 10000$ 個數中抽取 $n = 100$ 個樣本，得其母體平均值 90% 信賴區間為 (2.1 , 4.1)，如原始母體個數實為 100000，在其他結果不變時，則 90% 信賴區間的寬度將更寬。

Ans X

二、選擇題

1 ：母體平均值 μ 的 90% 信賴區間，其信賴係數為何？

A：0.1

B：0.05

C：0.9

D：0.95

Ans C

2 ：假設常態母體平均值 μ 的信賴區間為 $\bar{x} \pm 1.96\sigma/\sqrt{n}$，則其信賴水準 (Confidence Level) 為何？

A：97.5%

B：95%

C：0.975

D：0.95

Critical values of Standard Normal distribution

$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans B

3 ：假設你從母體重複實驗取樣 100 個隨機樣本，並用每一個隨機樣本建構母體平均值之 90% 信賴區間，試問這 100 個信賴區間中，你期望大約有多少個信賴區間包含真實的母體平均值？

A：100

B：95

C：90

D：10

Ans C

4 ：假設某母體比例之 95% 信賴區間為 (0.525, 0.725)，則抽樣樣本數約為多少？

A：120

B：64

C：91

D：180

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans C

5 ：假設在抽樣之前我們要決定樣本數，使得我們可以用樣本平均值來估計母體平均值，給定已知母體變異數、信賴水準及可容忍最大誤差，得知需抽樣 100 個樣本；今如可容忍最大誤差變成原本的一半，則我們變成需要抽樣多少樣本？

A：400

B：50

C：100

D：600

Ans A

6　：下列有關母體比率信賴區間之寬度，何者正確？

A：樣本大小為 49 時比樣本大小 36 時寬

B：95% 的信賴區間比 90% 的信賴區間寬

C：不受樣本數之影響

D：不受母體變異數大小之影響

Ans B

7　：隨機從常態母體中抽取 9 個樣本，得其樣本平均值及樣本變異數分別為 7 和 16，則母體平均值 95% 的信賴區間為

A：$7 \pm (1.96)\left(\frac{16}{\sqrt{9}}\right)$

B：$7 \pm (1.96)\left(\frac{4}{\sqrt{9}}\right)$

C：$7 \pm (1.645)\left(\frac{4}{\sqrt{9}}\right)$

D：$7 \pm (2.306)\left(\frac{4}{\sqrt{9}}\right)$

Critical values of Standard Normal distribution

$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Critical value of t distribution					
degree of freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
27	1.314	1.703	2.052	2.473	2.771
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
Normal	1.282	1.645	1.960	2.326	2.576

Ans D

8 ：某品管人員在管控色紙的平均長度，長度設定為 12 公分，已知色紙長度的標準差為 0.12 公分，品管人員如要有 99% 的信心，使得估計平均值的最大誤差在 0.03 公分內，則所需取樣的樣本數為何？

A：44

B：95

C：106

D：881

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans C

9 ：假設母體平均值的 90% 信賴區間為 (10, 20)，如此結果是根據樣本數 16 所計算出來，試問前提需有什麼假設成立，此信賴區間才是正確的？

A：母體資料結構近似於常態分配

B：母體資料結構近似於 t 分配

C：樣本平均值的抽樣分配是近似於常態分配

D：不需任何假設

Ans A

10 ：比較兩常態母體平均值時，如分別抽樣樣本數 n_1, n_2 皆是小樣本，且假設兩母體之標準差相等，則在使用 t 分配來建構信賴區間時，t 分配之自由度為

A：$n_1 - 1$

B：$n_2 - 1$

C：$n_1 + n_2 - 1$

D：$n_1 + n_2 - 2$

Ans D

11 ：檢定兩母體變異數是否相等時，我們需要何種假設？

A：兩母體資料皆為常態分配

B：兩母體資料皆為 t 分配

C：兩母體平均值相等

D：不需任何假設

Ans A

12 ：假設某母體比例之 95% 信賴區間為 (0.5, 0.7), 則母體比例之點估計值為何？

A：0.5

B：0.6

C：0.7

D：無法判斷

Ans B

13 ：某母體比例之區間估計為 (0.5, 0.7), 今天如果我們要重新取樣，使得我們有 99% 的信心，母體比例之點估計值誤差不超過 0.05，則需取樣多少樣本？

A：369

B：385

C：661

D：635

Critical values of Standard Normal distribution

$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans D

14 ：兩母體比例 p_1 , p_2 差之點估計式為 $\hat{p}_1 - \hat{p}_2$，其抽樣分配之平均值為

A：$\mu_1 - \mu_2$

B：0

C：$n_1 p_1 - n_2 p_2$

D：$p_1 - p_2$

Ans D

15 ：某校 1000 個女學生樣本中有 10% 抽菸，而 1500 個男同學樣本中則有 32% 抽菸，試問混合所有樣本後，其抽菸比率為

A：21%

B：23.2%

C：58%

D：42%

Ans B

16 ：某校 1000 個女學生樣本中有 10% 抽菸，而 1500 個男同學樣本中則有 32% 抽菸，試問女同學抽菸比率的 90% 信賴區間為何？

A： $0.1 \pm (1.96)(\sqrt{(0.1)(0.9)/1000}\,)$

B： $0.1 \pm (1.96)(\sqrt{(0.5)(0.5)/1500}\,)$

C： $0.32 \pm (1.96)(\sqrt{(0.32)(0.08)/1500}\,)$

D： $0.1 \pm (1.645)(\sqrt{(0.1)(0.9)/1000}\,)$

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans D

17 ：某校 1000 個女學生樣本中有 10% 抽菸，而 1500 個男同學樣本中則有 32% 抽菸，假設我們要檢定男生抽菸的比例和女生抽菸的比例不同，則檢定統計量為何？

A：$(0.32-0.1)/(\sqrt{(0.232)(0.768)(1/1000+1/1500})$

B：$(0.32-0.1)/(\sqrt{(0.1\times 0.9)/1000+(0.32\times 0.68)/1500})$

C：$(0.32-0.1)/(\sqrt{(0.232)(0.768)(1/2500})$

D：$(0.32-0.1)/(\sqrt{(0.1)(0.9)(1/2500})$

Ans A

18 ：在檢定非獨立成對資料平均差 μ_d 是否為 0 時，下列何者為 μ_d 之點估計值？

A：s_d

B：$\mu_1-\mu_2$

C：$\bar{d}$

D：0

Ans C

19 ：比較兩城市的平均薪資，得下列資料表

城市	樣本數	樣本平均值	樣本標準差	母體平均值
1	49	22250	700	μ_1
2	36	21370	600	μ_2

試問母體平均值差 $\mu_1-\mu_2$ 的點估計值為何？

A：3

B：880

C：100

D：440

Ans B

20 ：比較兩城市的平均薪資，得下列資料表

城市	樣本數	樣本平均值	樣本標準差	母體平均值
1	49	22250	700	μ_1
2	36	21370	600	μ_2

試問樣本平均值差 $\bar{X}_1 - \bar{X}_2$ 的標準差為何？

A：700

B：600

C：141

D：100

Ans C

21 ：比較兩城市的平均薪資，得下列資料表

城市	樣本數	樣本平均值	樣本標準差	母體平均值
1	49	22250	700	μ_1
2	36	21370	600	μ_2

試問 $\mu_1 - \mu_2$ 的 90% 信賴區間為何？

A：880 ± (1.96)(200)

B：880 ± (1.645)(141)

C：880 ± (1.645)(100)

D：880 ± (1.96)(141)

Critical values of Standard Normal distribution				
$Z_{0.10}$	$Z_{0.05}$	$Z_{0.025}$	$Z_{0.01}$	$Z_{0.005}$
1.282	1.645	1.960	2.326	2.576

Ans B

22 ：比較兩城市的平均薪資，得下列資料表

城市	樣本數	樣本平均值	樣本標準差	母體平均值
1	49	22250	700	μ_1
2	36	21370	600	μ_2

如要檢定 H_0: $\mu_1 - \mu_2 \geq 0$, H_a: $\mu_1 - \mu_2 < 0$, 則其檢定統計量約為何？

A：8.8

B：0.677

C：6.24

D：4.4

Ans C

23 ：比較兩城市的平均薪資，得下列資料表

城市	樣本數	樣本平均值	樣本標準差	母體平均值
1	49	22250	700	μ_1
2	36	21370	600	μ_2

如要檢定 H_0: $\mu_1 - \mu_2 \geq 0$, H_a: $\mu_1 - \mu_2 < 0$, 則其 P 值約為何？

A：1

B：0.8106

C：0.9861

D：0

Standard Normal table: probability for P(0 < Z <z)

z	0.00	*0.01*	0.02	*0.03*	0.04	*0.05*	0.06	*0.07*	0.08	*0.09*
0.0	0.0000	*0.0040*	0.0080	*0.0120*	0.0160	*0.0199*	0.0239	*0.0279*	0.0319	*0.0359*
0.1	0.0398	*0.0438*	0.0478	*0.0517*	0.0557	*0.0596*	0.0636	*0.0675*	0.0714	*0.0753*
0.3	0.1179	*0.1217*	0.1255	*0.1293*	0.1331	*0.1368*	0.1406	*0.1443*	0.1480	*0.1517*
0.4	0.1554	*0.1591*	0.1628	*0.1664*	0.1700	*0.1736*	0.1772	*0.1808*	0.1844	*0.1879*
0.5	0.1915	*0.1950*	0.1985	*0.2019*	0.2054	*0.2088*	0.2123	*0.2157*	0.2190	*0.2224*
0.6	0.2257	*0.2291*	0.2324	*0.2357*	0.2389	*0.2422*	0.2454	*0.2486*	0.2517	*0.2549*
0.8	0.2881	*0.2910*	0.2939	*0.2967*	0.2995	*0.3023*	0.3051	*0.3078*	0.3106	*0.3133*
0.9	0.3159	*0.3186*	0.3212	*0.3238*	0.3264	*0.3289*	0.3315	*0.3340*	0.3365	*0.3389*
1.0	0.3413	*0.3438*	0.3461	*0.3485*	0.3508	*0.3531*	0.3554	*0.3577*	0.3599	*0.3621*
1.1	0.3643	*0.3665*	0.3686	*0.3708*	0.3729	*0.3749*	0.3770	*0.3790*	0.3810	*0.3830*
1.2	0.3849	*0.3869*	0.3888	*0.3907*	0.3925	*0.3944*	0.3962	*0.3980*	0.3997	*0.4015*
1.4	0.4192	*0.4207*	0.4222	*0.4236*	0.4251	*0.4265*	0.4279	*0.4292*	0.4306	*0.4319*
1.5	0.4332	*0.4345*	0.4357	*0.4370*	0.4382	*0.4394*	0.4406	*0.4418*	0.4429	*0.4441*
1.6	0.4452	*0.4463*	0.4474	*0.4484*	0.4495	*0.4505*	0.4515	*0.4525*	0.4535	*0.4545*
1.7	0.4554	*0.4564*	0.4573	*0.4582*	0.4591	*0.4599*	0.4608	*0.4616*	0.4625	*0.4633*
1.9	0.4713	*0.4719*	0.4726	*0.4732*	0.4738	*0.4744*	0.4750	*0.4756*	0.4761	*0.4767*
2.0	0.4772	*0.4778*	0.4783	*0.4788*	0.4793	*0.4798*	0.4803	*0.4808*	0.4812	*0.4817*
2.2	0.4861	*0.4864*	0.4868	*0.4871*	0.4875	*0.4878*	0.4881	*0.4884*	0.4887	*0.4890*
2.3	0.4893	*0.4896*	0.4898	*0.4901*	0.4904	*0.4906*	0.4909	*0.4911*	0.4913	*0.4916*
2.4	0.4918	*0.4920*	0.4922	*0.4925*	0.4927	*0.4929*	0.4931	*0.4932*	0.4934	*0.4936*
2.5	0.4938	*0.4940*	0.4941	*0.4943*	0.4945	*0.4946*	0.4948	*0.4949*	0.4951	*0.4952*
2.6	0.4953	*0.4955*	0.4956	*0.4957*	0.4959	*0.4960*	0.4961	*0.4962*	0.4963	*0.4964*
2.7	0.4965	*0.4966*	0.4967	*0.4968*	0.4969	*0.4970*	0.4971	*0.4972*	0.4973	*0.4974*
2.9	0.4981	*0.4982*	0.4982	*0.4983*	0.4984	*0.4984*	0.4985	*0.4985*	0.4986	*0.4986*
3.0	0.4987	*0.4987*	0.4987	*0.4988*	0.4988	*0.4989*	0.4989	*0.4989*	0.4990	*0.4990*
3.1	0.4990	*0.4991*	0.4991	*0.4991*	0.4992	*0.4992*	0.4992	*0.4992*	0.4993	*0.4993*
3.3	0.4995	*0.4995*	0.4995	*0.4996*	0.4996	*0.4996*	0.4996	*0.4996*	0.4996	*0.4997*
3.4	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4997*	0.4997	*0.4998*

Ans A

三、複選題

1　：建構母體平均值之信賴區間，其信賴區間之寬 (長) 度，受那些因素的影響？

A：樣本大小

B：信賴水準

C：母體平均值之大小

D：母體標準差

E：樣本平均值之大小

Ans A, B, D

2 ：下列有關母體比率差的點估計 $\hat{p}_1 - \hat{p}_2$ 之抽樣分配，何者為真？

A：不管樣本數大小，皆近似常態分配

B：其平均值為 $p_1 - p_2$

C：其平均值為 $\mu_1 - \mu_2$

D：其標準差為 $p_1q_1/n_1 + p_2q_2/n_2$

E：其變異數為 $p_1q_1/n_1 + p_2q_2/n_2$

Ans B, E

Word

- Word 2013文件編輯技巧
- 建立圖文並茂的檔案
- Word版面的設定
- 索引目錄的製作
- 頁首頁尾的建立
- 文件的追蹤與修訂

Excel

- 工作表的建立
- 函數和公式的使用知識
- 建立和編輯圖表
- 表格的排序與統計
- 樞紐分析表的建立
- 資料規劃與分析

PowerPoint

- PowerPoint 2013
- 簡報編輯知識
- 完全認識投影片的母片
- 建立含特效的簡報
- 投影片放映的完整知識

Office2013 Word、Excel、PowerPoint 國際認證

Silicon Stone Education的Office Word、Excel、PowerPoint證照，是您可以通行全球，跟隨您一輩子的重要能力證明，類似於英國劍橋大學的雅思(IELTS)，美國ETS的TOEFL，未來將是通行於全球的國際權威認證之一。

計分方式 Exam Scoring

本考科共60題，是非題有20題，選擇題有35題和複選題5題，每題有20分，滿分是1200分，840分及格分數。

考試會參考當次考生分數及難易程度做動態式調整。

證照資料 License Information

通過測驗後，將寄出姓名為中英文之合格證書，至考試中心領取，額外申請中文或英文證書，須支付工本費500元

證照樣式 License Sample

QR Code
辨識認證資格

- ▲ 增強個人履歷
- ▲ 提供就業或研究所推甄有力證明
- ▲ 突顯專業技能，超越同儕
- ▲ 學習第二專長，創造更多就業機會

- ▲ 產業結合密切
- ▲ 具備工作應用之廣度及深度
- ▲ 增加額外收入的最短捷徑
- ▲ 提升開發學習專業領域技能

台灣獨家授權代理經銷商

上奇資訊 服務專線：02-2562-7969 傳真專線：02-2562-5269
http://www.grandtech.info/

佳魁資訊 服務專線：02-2562-7756 傳真專線: 02-2562-7716
http://www.topteam.cc /

Join Now

104 台北市中山區南京東路二段98號8樓之1

Lausanne
Project

www.onlinetest.com.tw

BEST QUALITY

CERTIFIED LOUNGE-BAR PROFESSIONAL

精品咖啡師證照

咖啡的地理與歷史
咖啡豆挑選和加工
研磨和沖泡方式
咖啡飲料的種類與特色
對人體的好處
吧檯清潔，安全與衛生

時尚調酒師證照

酒的分類與釀製
認識紅葡萄酒和白葡萄酒
雞尾酒基礎與實務
雞尾酒器具認識
酒與食物之搭配
吧檯清潔，安全與衛生

茶藝師證照

茶的栽種環境與採摘
茶的發酵識別分類
茶的種類與製作方法
喝茶的益處
吧檯清潔，安全與衛生

計分方式 Exam Scoring

本考科共60題，是非題30題，選擇題30題，每題10分，滿分是600分，420分為及格分數。
考試費用：新台幣1,980元

證照資料 License Information

通過測驗後，將寄出姓名為中英文之合格證書，至考試中心領取，額外申請中文或英文證書，須支付工本費500元 中文或英文證書，須支付工本費500元

證照樣式 License sample

QR Code
辨識認證資格

考取上述3項證照，Silicon Stone主動授與吧檯專業技師證照

Silicon Stone則主動授與吧檯專業技師(Certified Lounge-Bar Professional)證照
本公司將寄出姓名為中英文之合格證書，請至考試中心領取。
我們深信Lausanne吧檯專業技師(Certified Lounge-Bar Professional)證照，是您可以通行全球，跟隨您一輩子的重要能力證明，未來將是通行於全球餐旅界的國際權威認證之一。

台灣獨家授權代理經銷商

上奇資訊 服務專線：02-2562-7969 傳真專線：02-2562-5269
http://www.grandtech.info/

佳魁資訊 服務專線：02-2562-7756 傳真專線: 02-2562-7716
http://www.topteam.cc /

Join Now

104 台北市中山區南京東路二段98號8樓之1

Computer Knowledge Today

國際認證

瞬息萬變的電腦知識掌握最新知識

滿足未來求學或就業的需求

Computer
電腦基本觀念
電腦資料表示法
數字系統

Equipment
基本電腦硬體知識
基本電腦軟體知識

Website
網頁瀏覽知識
電子郵件知識

Security
資訊安全
網路著作權

Silicon Stone

Computer Knowledge Today
由散佈於全球計算機領域的教育專家與Silicon Stone Education合作所推廣的國際證照，是您可以通行全球，跟隨您一輩子的重要能力證明，相信未來將可類似於英國劍橋大學的雅思(IELTS)，美國ETS的TOEFL，通行於全球。

計分方式 Exam Scoring
本考科共50題，是非題10題
選擇題35題，複選題有5題
每題有20分，滿分是1000分
700為及格分數。

證照資料 License Information
通過測驗後，將寄出姓名為中文之合格證書，至考試中心領取，額外申請中文或英文證書，須支付工本費500元

證照樣式 License Sample


QR Code
辯識認證資格

- ▲ 增強個人履歷
- ▲ 提供就業或研究所推甄有力證明
- ▲ 凸現專業技能，超越同儕
- ▲ 學習第二專長，創造更多就業機會

- ▲ 產業結合密切
- ▲ 俱備工作應用之廣度及深度
- ▲ 增加額外收入的最短捷徑
- ▲ 提升開發學習專業領域技能

台灣獨家授權代理經銷商

上奇資訊 服務專線：02-2562-7969 傳真專線：02-2562-5269
http://www.grandtech.info/

佳魁資訊 服務專線：02-2562-7756 傳真專線: 02-2562-7716
http://www.topteam.cc /

Join Now

104 台北市中山區南京東路二段98號8樓之1

Silicon Stone

全球唯一攝影系列認證

The world's only photography certification series

WELCOME

Photography

關於 Silicon Stone

Silicon Stone Education成立了，我們謹慎地建立線上測驗系統和考科資料庫。每個考科皆是由這個領域的多個專家，反覆推敲，討論而成，期待可對所有參與者有幫助。

我們認真看待所從事的工作，深知證照的取得是一份榮譽，我保證這將會是全球首創質感最好的證照，這份證照與榮譽將伴隨您一輩子。另外在證照內，本公司也獨創可放考生的照片，並加上QR code，考生，學校或企業均可隨時進本公司查詢真偽。

Photography

Fundamental concepts

攝影基本概念與常識
各式鏡頭的使用時機與效果

Aperture

Shutter and exposure

光圈、快門和曝光知識
不同情況下各種攝影參數設置

Composition

Actual usage

構圖理論與實際使用情況
後置軟體的基礎掌握

計分方式 Exam Scoring

本考科共60題，是非題20題
選擇題35題，複選題有5題
每題有20分，滿分是1200分
840為及格分數。

證照資料 License Information

通過測驗後，將寄出姓名為中文之合格證書，至考試中心領取，額外申請中文或英文證書，須支付工本費500元

證照樣式 License sample

QR Code
辯識認證資格

行銷、平面設計、廣告業、網頁設計、平面出版、微電影、遊戲製作、動畫、網拍
電子商務、App、虛擬攝影棚、互動開發、後製特效、電影、商業攝影、擴增實境

台灣獨家授權代理經銷商

上奇資訊 服務專線 : 02-2562-7969 傳真專線 : 02-2562-5269
http://www.grandtech.info/

佳魁資訊 服務專線 : 02-2562-7756 傳真專線: 02-2562-7716
http://www.topteam.cc /

Join Now

104 台北市中山區南京路二段98號8樓之1

上奇資訊股份有限公司

Grandtech Information Co., Ltd.