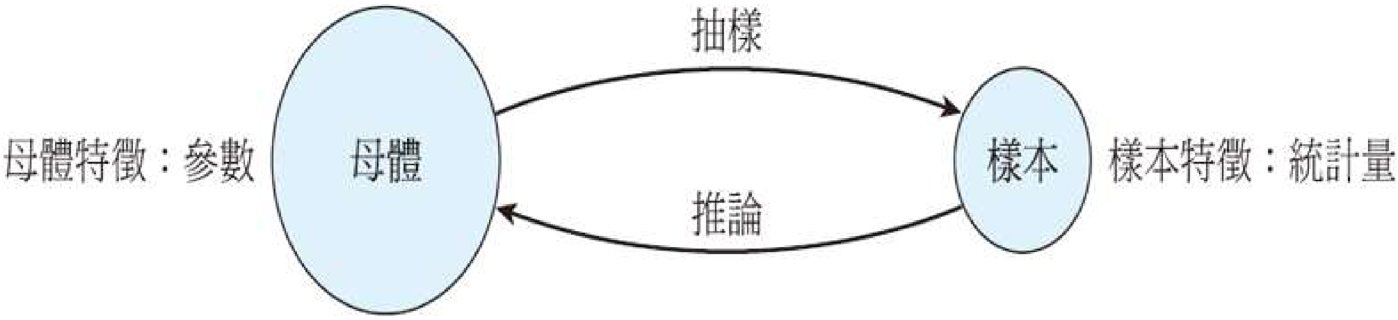
8 抽樣分配

抽樣分配是指樣本統計量的機率分配。

8-1 抽樣

一葉知秋：瞭解現象（樣本），推論真相（母體）



[參考影片：抽樣分配]

8-1-1 什麼是抽樣

1. 抽樣的定義

抽樣(sampling)是指自一母體取得樣本的程序，其目的在於以經濟有效的方法抽出一組具有代表性的樣本，以利用樣本對母體的特性進行推。論

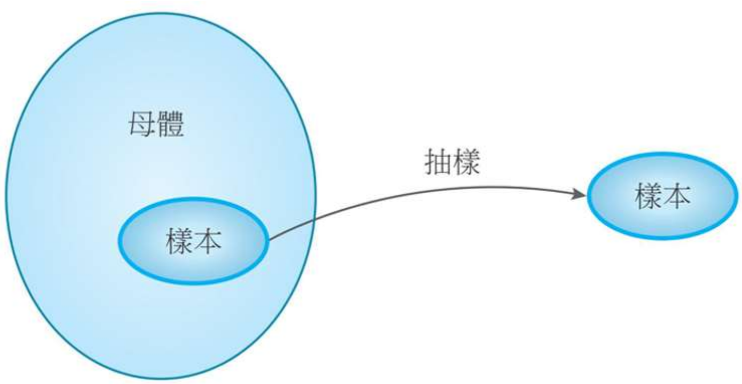


圖 1.3　抽樣示意圖

2. 為什麼要抽樣？

例如：要如何知道綠豆湯夠不夠甜？濃稠度適不適宜？

- 有人會喝完整鍋綠豆湯給答案嗎？

- 拿個湯匙先攪拌均勻，然後舀一杓嚐看看，以便知道綠豆湯整鍋的甜度與濃度。

=> 母體資料不可得，或是獲得母體資料代價太高時，先由母體取得代表性樣本，再由樣本統計量推論母體參數。

- 取得「代表性樣本」是關鍵。
- 當母體與樣本特性相當時，樣本特性就可以代表母體特性，亦即可以利用樣本統計量推論母體參數。

8-1-2 如何抽樣

1. 抽樣的方法

一般常用的抽樣方法可分為隨機抽樣(random sampling)與非隨機抽樣(non-random sampling)，前者所抽出的樣本具有隨機性(randomness)，故可進一步利用統計方法（機率）對樣本的特性加以探討；而後者則無此一功能。

- **非隨機抽樣法：**沒有按照機率的原則或是由研究者主觀認定所進行的抽樣方法。常用的有便利抽樣法、立意抽樣法、雪球抽樣法等。

- **隨機抽樣法**: 依照機率的原則，抽取隨機樣本的過程稱為隨機抽樣。

假設由一母體抽出n筆資料， 以隨機變數$X_1, X_2, ..., X_n$表示。 如果滿足(1)這n個隨機變數皆彼此相互獨立， 以及(2)這n個隨機變數皆抽自同一母體這兩個條件， 則稱這n筆資料，$X_1, X_2, ..., X_n$， 為由此母體抽出的一組隨機樣本。

2. 隨機抽樣的方法分類

- **簡單隨機抽樣法(simple random sampling)**
  - 簡單隨機抽樣法是最基礎的機率抽樣方法，「使母體中的每一元素被抽中的機率都相同」。
  - 常見的執行方式有抽籤、使用亂數表或使用電腦產生亂數。

例如：抽籤法

對一個含有N 個對象的有限母體，假設每一個對象被抽出的機會均相同，則可以抽籤法來獲得隨機樣本，此方法的抽取步驟如下：

```
- 將母體由 1號編到 N 號。
- 製作 N 支籤，上面編上 1 至 N 號。
- 將 N 支籤均勻的放置桶內。
- 由籤桶中抽出 n 支籤。
- 所抽出之 n 支籤的號碼即為所抽取的樣本。
```

- **分層隨機抽樣(stratified sampling)**： 分層隨機抽樣的取樣方法包括三個步驟， (1)將母群分成幾個階層；(2)對每個階層實施隨機抽樣；(3)估計母群之均值。

例如將台灣地區359個鄉鎮區分成七個選樣層，然後每層各自以隨機抽取方式抽100人。

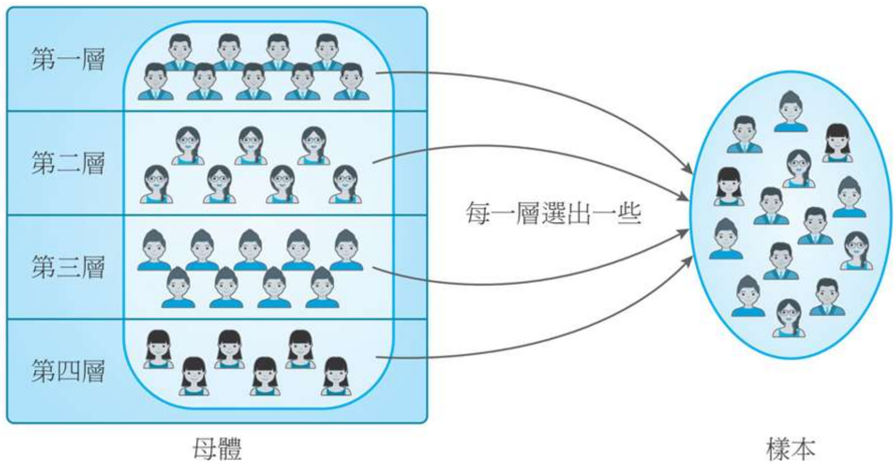


圖 7.1　分層隨機抽樣示意圖

- **系統抽樣(systematic sampling)**： 在母體中依固定的間隔選取樣本，又稱等距抽樣。例如管理學院學號能被7除進的同學。

例如：假定母體中有100個單元，編號由1到100，則一個簡單的系統抽樣進行的方式可以是先由0~9中以簡單隨機的方式任意選取一個數字，假定被選取的是3，則由3開始，每10個間隔就選取一個單元，則產生之樣本為3, 13, 23, 33, 43, 53, 63, 73, 83, 93系統抽樣的優點是只要抽出第一個種子號碼，就可依間隔數依序列出所有樣本數，所以此種抽樣法比簡單隨機抽樣法來的方便。

1	2	3
4	5	6
7	8	9
10	11	12
13	14	15

圖 7.2　系統抽樣示意圖

- **集群抽樣(cluster sampling)**： 將母體區分為許多不同的集體，然後抽取少數集體當作樣本集體，抽中的集體全部調查。例如大一到大四各抽一班，抽中的班級全班同學接受調查。

集群抽樣是將整個母體依其特性分成若干群集(cluster)，在進行抽樣時所選取的是群集，然後抽取少群集體當作樣本集體，抽中的群集全部調查。例如大一到大四各抽一班，抽中的班級全班同學接受調查。當然我們亦可在所選取之集群中再進行隨機抽樣以選取樣本，通常在這種情況下稱之為二階段(two-stage)抽樣。

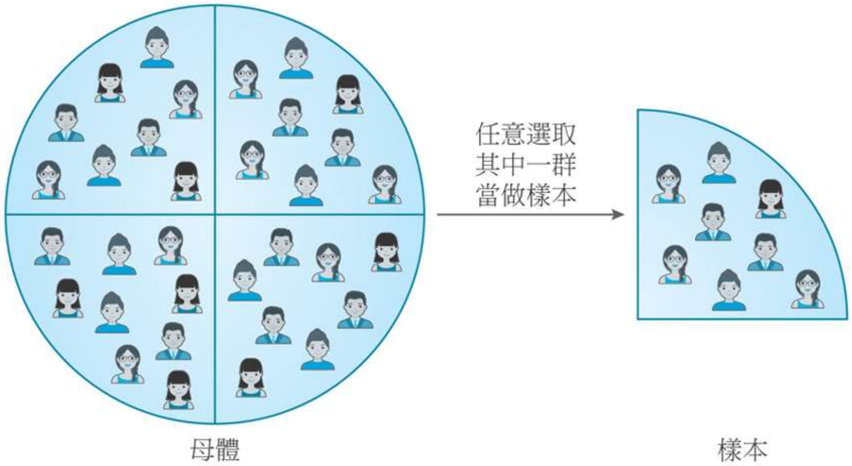


圖 7.3 集群抽樣法示意圖

=> 例如要了解世新大學學生學測的平均成績，不同的抽樣方法，抽出來的樣本不同，答案也會跟著不一樣。通常，要選擇一個合適的機率抽樣方法，來降低因為抽樣方法的不同所造成的估計誤差。

8-1-3 抽樣誤差

抽樣誤差 = 樣本統計量 - 母體參數 = 系統誤差 ＋ 隨機誤差

以飛鏢射標為例

- 系統誤差是指投擲的飛鏢都往同一個方向偏離靶心的偏差
  - 現在的民調若是用電話進行訪問，受訪者多為老年人，這就是系統誤差。
  - 若是用網路調查進行訪問，受訪者多為年輕人、高學歷者，這就是系統誤差。
- 隨機誤差是指投擲的飛鏢在靶上分散程度的變異性。
- 好的抽樣設計希望能得到具備有低偏差和低變異的特性。

樣本提供的資訊是母體的縮影

- 我們將樣本帶入統計量的公式，計算出樣本統計量，藉由樣本統計量來估計母體參數。然而每次抽樣時，抽取的樣本皆不同，計算出來的樣本統計量跟著改變，因此會造成統計量跟母體參數之間有一些誤差。
- 所謂抽樣誤差(sampling bias)就是母體參數和統計量的差距，而抽樣誤差包括系統誤差(systematic bias)與隨機誤差(random bias)，系統誤差會導致統計量有所偏差，隨機誤差會造成統計量分散。

降低偏差的方法

- **使用隨機抽樣得到的樣本統計量來估計母體參數。** 理論上說如果做很多次會得到很多個統計值，把這些統計值平均起來就會很接近真實的結果。
- **減少變異性的方法是取得大一點的樣本。**也就是說，只要樣本取得夠大，雖然每次做出來的結果會不相同，但是差距不會很大，樣本的變異性就變小了。

8-2 樣本統計量的機率分配

母體隨機變數有自己的機率分配；將抽出來的樣本，利用統計量的公式計算，也有樣本統計量隨機變數的機率分配；二種機率分配的隨機變數定義不同，機率分配也不一樣。

8-2-1 什麼是抽樣分配

1. 樣本統計量的機率分配稱為抽樣分配(Sampling distribution)。

每次抽樣，抽出來的樣本可以代入不同的統計量公式中，計算出這個樣本的平均數、變異數等等統計量數。如果我們多抽幾次，就會發現發每次隨機抽出來的樣本都不一樣，帶入統計量公式所計算出來的數值也不同。因此，樣本統計量本身就是一個隨機變數，可以形成一個機率分配，常見的有樣本平均數的抽樣分配、樣本比例的抽樣分配，以及樣本變異數的抽樣分配。

2. 隨機樣本的特性

- **樣本變異性**：每次隨機抽樣的樣本都不同。
- **統計量隨機性**：將每次抽出來的樣本帶入統計量公式，計算出來的樣本平均數、樣本比例、樣本變異數等數值都不同，因此每一個樣本統計量都是一個隨機變數，可以各自形成樣本統計量的機率分配(抽樣分配)。
- **抽樣誤差**：透過樣本計算出來的統計量數和母體參數的真正值不同、會有些差距，這個樣本統計量與母體參數的差距稱為抽樣誤差，又名估計誤差。
    - 估計值 = 參數 ＋ 抽樣誤差
    - 例如樣本平均數與母體平均數的抽樣誤差為，$\bar{X} - \mu$。
    - 造成抽樣誤差的原因很多，例如樣本統計量、樣本大小、測量誤差（問卷設計如敏感性問題等、穿著大衣量體重）、抽樣方法、抽樣框架誤差（電訪是手機、戶機，網路用戶）等。

8-2-2 抽樣分配的求取

例題 7.1

設有一母體由 $\{1,2,3,4\}$ 所組成，今隨機抽出兩個數，即樣本大小為 $n=2$ 的一組樣本。抽取方式為一次抽一個，以取出後放回的方式進行抽樣。令統計量為樣本平均數 $\bar{X}=(X_1+X_2)/2$，試求 $\bar{X}$ 的抽樣分配及其抽樣分配的期望值與變異數。

解 利用三個步驟解析 $\bar{X}$ 的抽樣分配：

步驟 1：討論母體的機率分配

母體由 $\{1,2,3,4\}$ 所組成，母體大小為 $N=4$，假設母體中每個元素都有相同機率被抽中，則其機率分配如下：

x	$P(X=x)$
1	1/4
2	1/4
3	1/4
4	1/4

母體參數為

$$\mu=1\times\frac{1}{4}+2\times\frac{1}{4}+3\times\frac{1}{4}+4\times\frac{1}{4}=\frac{10}{4}=2.5$$

$$\sigma^2=(1-2.5)^2\times\frac{1}{4}+(2-2.5)^2\times\frac{1}{4}+(3-2.5)^2\times\frac{1}{4}+(4-2.5)^2\times\frac{1}{4}=\frac{5}{4}=1.25$$

步驟 2：討論抽樣結果

隨機抽取 $n=2$ 為一組樣本，共有 $N^n=4^2=16$ 組可能樣本，每一組樣本發生的機率均為 1/16，如下表所示。

樣本	$\bar{X}$ 的值	發生機率	樣本	$\bar{X}$ 的值	發生機率
{1 , 1}	1	1/16	{3 , 1}	2	1/16
{1 , 2}	1.5	1/16	{3 , 2}	2.5	1/16
{1 , 3}	2	1/16	{3 , 3}	3	1/16
{1 , 4}	2.5	1/16	{3 , 4}	3.5	1/16
{2 , 1}	1.5	1/16	{4 , 1}	2.5	1/16
{2 , 2}	2	1/16	{4 , 2}	3	1/16
{2 , 3}	2.5	1/16	{4 , 3}	3.5	1/16
{2 , 4}	3	1/16	{4 , 4}	4	1/16

步驟 3： 討論抽樣分配

令樣本統計量：$\bar{X}=(X_1+X_2)/2$，即以 $n=2$ 的樣本的平均數為樣本統計量。

將上表的所有可能的樣本平均數及其發生的機率加以整理，即可得出樣本平均數 $\bar{X}$ 的抽樣分配。

$\bar{x}$	$P(\bar{X}=\bar{x})$
1	1/16
1.5	2/16
2	3/16
2.5	4/16
3	3/16
3.5	2/16
4	1/16

樣本平均數 $\bar{X}$ 的抽樣分配的期望值與變異數

$$E(\bar{X})=1\times\frac{1}{16}+1.5\times\frac{2}{16}+2\times\frac{3}{16}+2.5\times\frac{4}{16}+3\times\frac{3}{16}+3.5\times\frac{2}{16}+4\times\frac{1}{16}=\frac{40}{16}=2.5=\mu$$

$$\begin{aligned}V(\bar{X})&=(1-2.5)^2\times\frac{1}{16}+(1.5-2.5)^2\times\frac{2}{16}+(2-2.5)^2\times\frac{3}{16}+(2.5-2.5)^2\times\frac{4}{16}\\&\quad+(3-2.5)^2\times\frac{3}{16}+(3.5-2.5)^2\times\frac{2}{16}+(4-2.5)^2\times\frac{1}{16}\\&=\frac{10}{16}=0.625=\frac{\sigma^2}{2}\end{aligned}$$

=> 我們發現

- 母體機率分配和樣本統計量（平均數）的機率分配
  - 隨機變數定義不同：X與$\bar{X}$
  - 機率分配不同：$(X, f(X))$與$(\bar{X}, f(\bar{X}))$
- 求特性
  - 期望值：$E(X)$與$E(\bar{X})$相同
  - 標準差：$Var(x)$與$Var(\bar{X})$不同。

8-2-3 影響抽樣分配的因素

- 母體機率分配不同，抽樣分配不同 = > 抽樣分配是什麼
- 樣本數不同，抽樣分配不同 = > 中央極限定理
- 樣本統計量不同，抽樣分配不同 = > 好的點估計量

【加分題】擲骰子的抽樣分配

(1)（抽樣分配）投擲一公正骰子2次，以X_1 表示第一次所出現的結果，X_2表示第二次所出現的結果，令$\bar{X} = \frac{X_1+X_2}{2}$，試求$\bar{X}$的機率分配。

(2)（母體機率分配不同）投擲一骰子2次，骰子出現點數為$1, 2, 3, 4, 5, 6$的機率分別為$0.1, 0.2, 0.2, 0.2, 0.2, 0.1$。若以$X_1$ 表示第一次所出現的結果，X_2表示第二次所出現的結果，令$\bar{X} = \frac{X_1+X_2}{2}$，試求$\bar{X}$的機率分配。

(3)（樣本數不同）投擲一公正骰子3次，以X_1 表示第一次所出現的結果，X_2表示第二次所出現的結果，X_3表示第三次所出現的結果，令$\bar{X} = \frac{X_1+X_2+X_2}{3}$，試求$\bar{X}$的機率分配。

(4)（統計量不同）投擲一公正骰子2次，以X_1 表示第一次所出現的結果，X_2表示第二次所出現的結果，令$\bar{X} = \frac{2X_1+1X_2}{3}$，試求$\bar{X}$的機率分配。

=> 上面四題計算出來的樣本平均數的機率分配均不同；抽樣分配不同，做出的統計推論結果也不一定會相同。

Part One: 單一母體

8-3 樣本平均數的抽樣分配

8-3-1 基本特性

1. 樣本平均數的抽樣分配示意圖

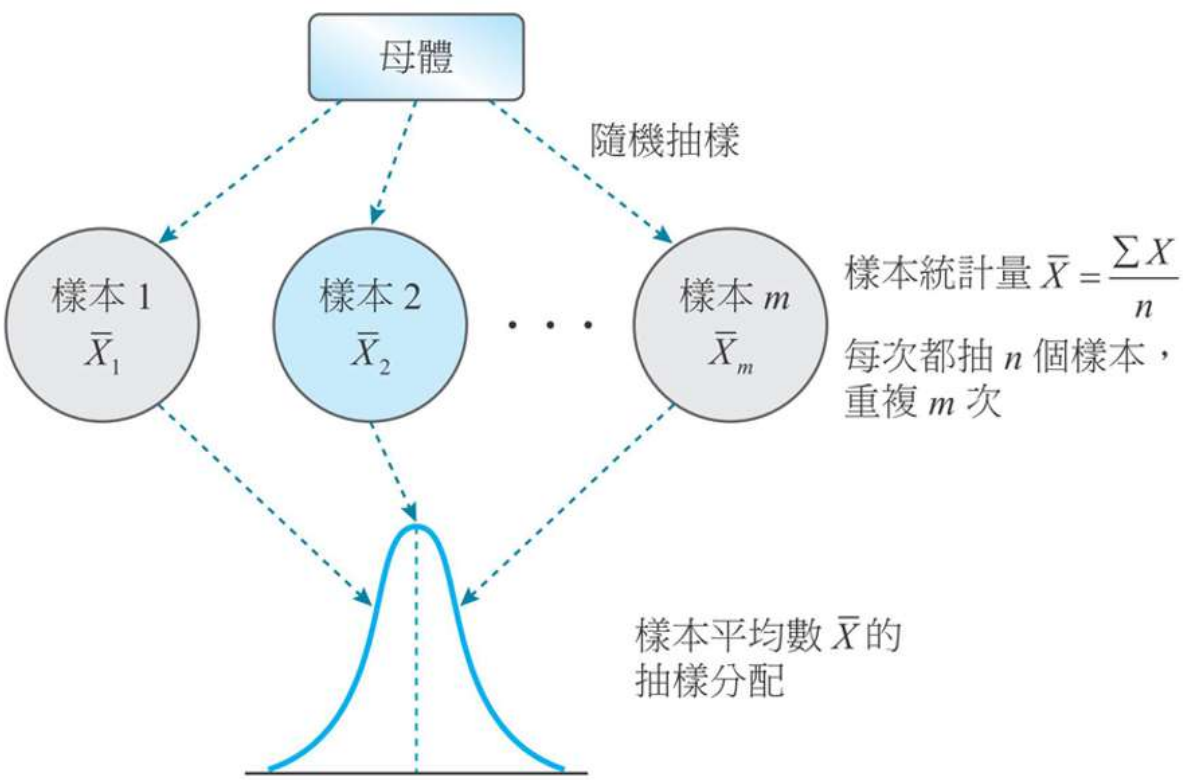


圖 7.4 樣本平均數的抽樣分配示意圖

2. 樣本平均數抽樣分配

1. 定義

假設母體為任意分配，其期望值為μ，變異數為σ^2。從此母體中，隨機且獨立的抽出大小為n的一組隨機樣本$X_1, X_2, \ldots, X_n$，透過樣本平均數這個統計量公式$\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n}$，可推得樣本平均數統計量這個隨機變數所形成的機率分配，稱為樣本平均數抽樣分配。

(2)求特性：期望值與變異數

- 樣本平均數抽樣分配的期望值與變異數：
  - $\mu_{\bar{X}} = E(\bar{X}) = \mu$
  - $\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$
- 樣本平均數$\bar{X}$的標準差稱為**標準誤(standard error)**
  - $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
    - 當樣本數固定時，母體分配的分散程度σ 直接影響$\bar{X}$分配分散程度$\sigma_{\bar{X}}$ 的大小。
    - 當σ固定時，樣本數 n 增大，$\bar{X}$ 分配的分散程度$\sigma_{\bar{X}}$ 會隨之減少。

Proof:

期望值公式: $E(X) = \sum x_i f(X = x_i) = \mu$

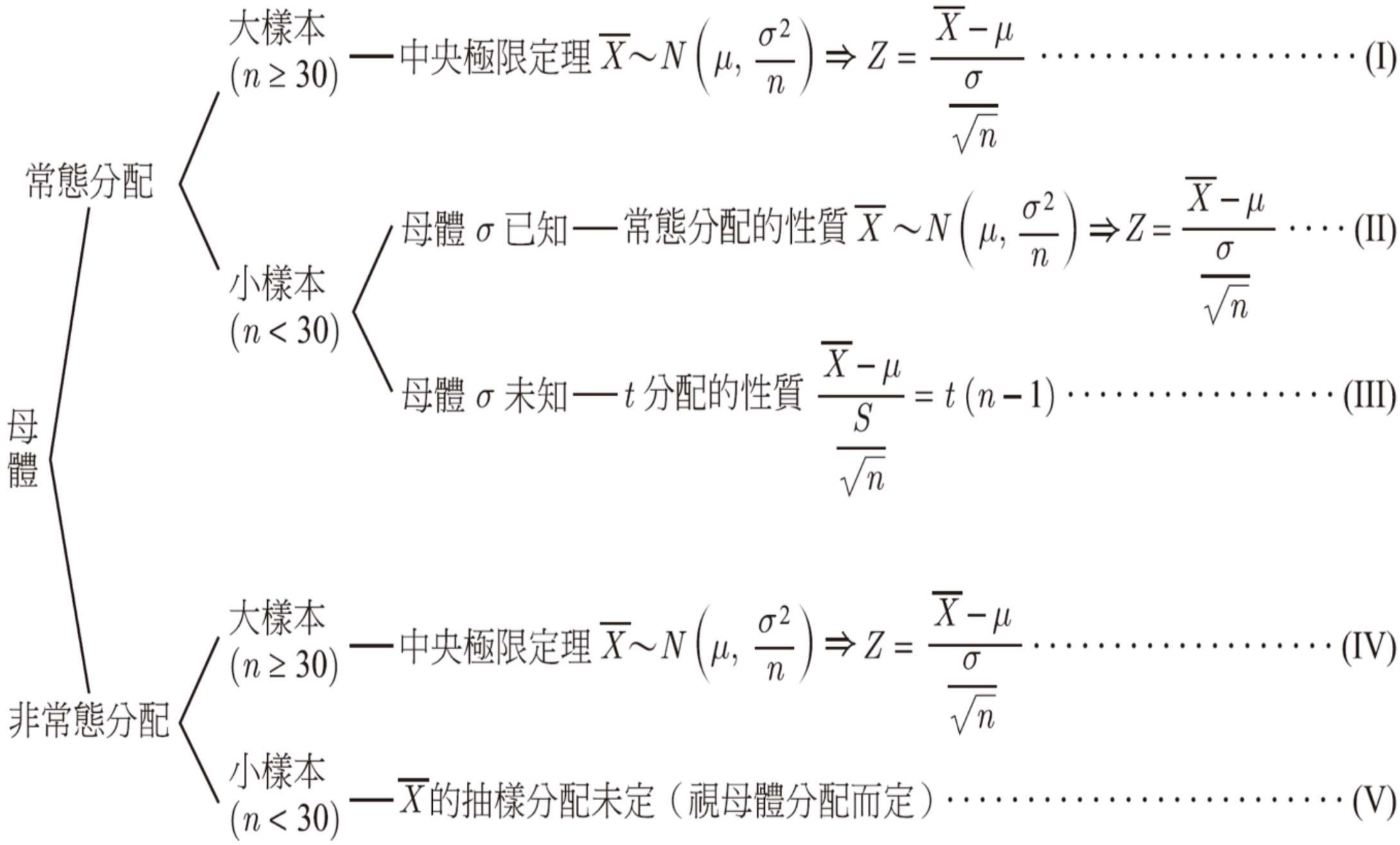
$\mu_{\bar{X}} = E(\bar{X}) = E(\frac{X_1+\ldots+X_n}{n}) = \frac{E(X_1)+E(X_2)+\ldots+E(X_n)}{n} = \frac{nE(X)}{n} = E(X) = \mu$

變異數公式：$V(X) = \sum(x_i - \mu)^2 f(X = x_i) = \sigma^2$

$\sigma_{\bar{X}}^2 = V(\bar{X}) = V(\frac{X_1+\ldots+X_n}{n}) = V(\frac{X_1}{n}) + \ldots + V(\frac{X_n}{n}) = \frac{V(X)}{n^2} + \ldots + \frac{V(X)}{n^2} = \frac{V(X)}{n} = \frac{\sigma^2}{n}$

(3) 算機率：要知道確切的樣本平均數抽樣分配是什麼才可以算機率

母體是否為常態、樣本的大小、變異數是否已知三個條件決定抽樣分配是使用那一個機率分配。



例：我們想知道全台灣20歲以上民眾(母體)的平均身高。

如果隨機抽出100人，則可以得到這100人的平均身高是144公分；理論上如果無窮次的抽下去，就會得到無窮個身高平均數。

- 這些所有「身高平均數」的平均數，就會是母體的身高平均數。
- 這些所有身高平均數的變異數，就會是母體的身高平均數除以樣本大小。
- 如果母體的身高是常態分配，那麼這些身高平均數也是常態分配
- 如果母體的身高不是常態分配，那麼這些身高平均數會近似常態分配

8-3-2 熟能生巧

假設母體為任何分配，其期望值為μ，變異數為σ^2。 從此母體中，獨立且隨機的抽出大小為n的一組隨機樣本， $X_1, ..., X_2$，帶入樣本平均數的公式$\bar{X} = \frac{X_1 + ... + X_n}{n}$， 則樣本平均數抽樣分配的期望值與變異數分別為 $E(\bar{X}) = \mu$ 和$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$。

- 樣本平均數的標準差為其變異數開根號，稱為標準誤$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$。

=> 若母體分配為常態分配$N(\mu, \sigma^2)$，則樣本平均數$\bar{X}$ 的抽樣分配為常態分配$N(\mu, \sigma^2/n)$。

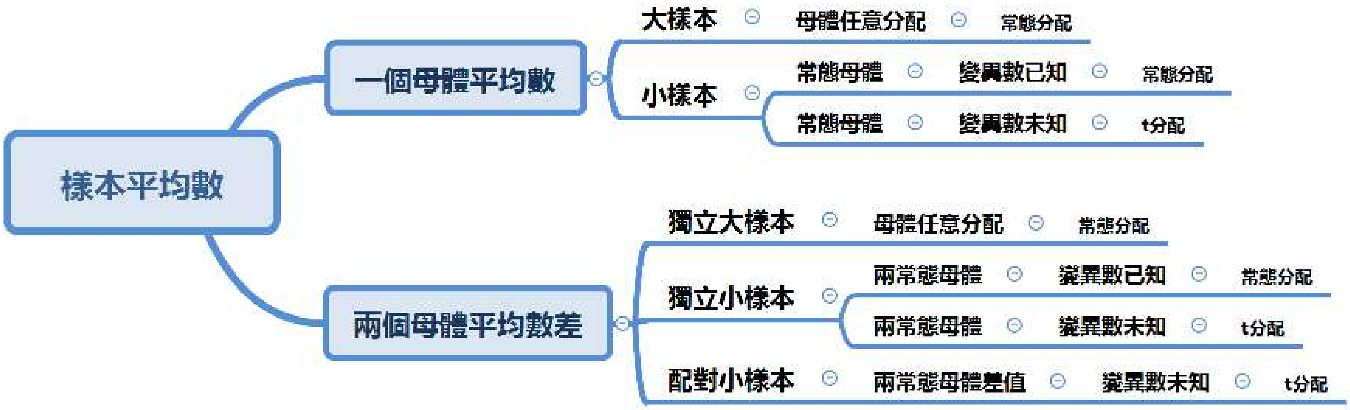
計算樣本平均數$\bar{X}$的機率問題時，

- 變異數σ^2已知，將$\bar{X}$標準化， $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$；
- 變異數σ^2未知，將$\bar{X}$轉成t分配， $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$。

=> 若母體為任意分配，在大樣本時，樣本平均數$\bar{X}$的抽樣分配 近似於常態分配$N(\mu, \sigma^2/n)$。

計算樣本平均數$\bar{X}$的機率分配時

- 變異數σ^2已知，將$\bar{X}$標準化，$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$。
- 變異數σ^2未知，將$\bar{X}$轉成t分配，$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$，而在大樣本時，$t \rightarrow N(0,1)$，仍查標準常態分配表。



(用t分配的只有母體是常態分配、小樣本、母體標準差未知這一種情況，其他除非是不知道機率分配是什麼，知道的就一定常態分配。)

例題 7.2

鮪魚罐頭製造商在罐頭標示固形物 140 公克。實際上，**固形物的重量是平均數 140 公克和標準差 4 公克的常態分配**。假設你在賣場購買 9 罐的鮪魚罐頭，量測固形物的重量為樣本。試求：

(1) 樣本平均數的抽樣分配。

(2) 樣本平均數少於 138 公克的機率。

解 (1) 母體為鮪魚罐頭的固形物重量，服從常態分配 $N(140, 4^2)$。

在賣場購買 9 罐的鮪魚罐頭為樣本，得罐頭的固形物重量 $X_1, \cdots, X_9$ 為獨立且有相同的機率分配 $N(140, 4^2)$。

樣本平均數 $\bar{X} = \dfrac{X_1 + \cdots + X_9}{9}$ 的期望值與變異數分別為

$$E(\bar{X}) = 140$$

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{4^2}{9}$$

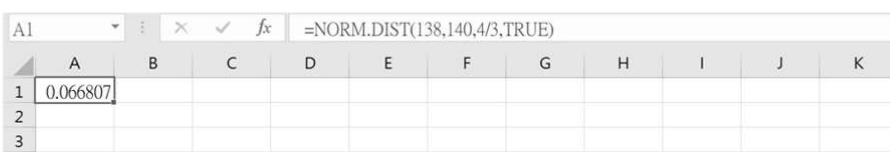
樣本平均數 $\bar{X}$ 的抽樣分配為常態分配 $N(140, 4^2/9)$

(2) $\bar{X} \sim N(140, 4^2/9)$

將 $\bar{X}$ 標準化成標準常態分配的隨機變數 Z，

$$Z = \frac{\bar{X} - 140}{4/\sqrt{9}} \sim N(0, 1)$$

$$P(\bar{X} < 138) = P\left(\frac{\bar{X} - 140}{\frac{4}{\sqrt{9}}} < \frac{138 - 140}{\frac{4}{\sqrt{9}}}\right) = P(Z < -1.5) = 0.0668$$



Excel 2016 函式：=NORM.DIST(138,140,4/3,TRUE)

由此，鮪魚罐頭的固形物重量的樣本平均數少於 138 公克的機率大約為 0.07，顯示鮪魚罐頭的固形物重量的樣本平均數有很高的可能性是不少於 138 公克。

例題 7.3

鮪魚罐頭製造商在罐頭標示固形物 140 公克。實際上，**固形物的重量是平均數 140 公克和標準差 σ 未知的常態分配**。假設你在賣場購買 9 罐的鮪魚罐頭，量測固形物的重量為樣本，得一組樣本標準差 s 為 3 公克。試求：在賣場購買 9 罐的鮪魚罐頭的樣本平均數少於 138 公克的機率。

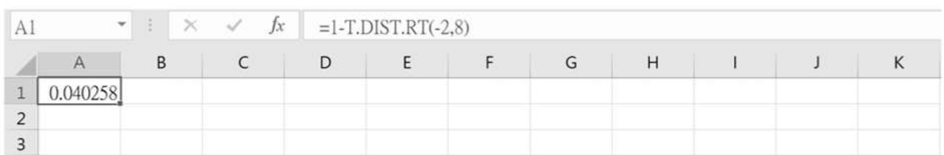
解 母體為鮪魚罐頭的固形物重量，服從常態分配 $N(140, \sigma^2)$。

在賣場購買 9 罐的鮪魚罐頭為樣本，得罐頭的固形物重量 $X_1, \cdots, X_9$ 為獨立且有相同的機率分配 $N(140, \sigma^2)$，母體變異數 σ^2 未知。

用樣本標準差 $s = 3$ 估計母體標準差 σ。$n-1 = 9-1 = 8$。

$$t = \frac{\bar{X} - 140}{3/\sqrt{9}} \sim t_8$$

$$P(\bar{X} < 138) = P\left(\frac{\bar{X} - 140}{\frac{3}{\sqrt{9}}} < \frac{138 - 140}{\frac{3}{\sqrt{9}}}\right) = P(t < -2) = 0.040258$$



Excel 2016 函式：=1-T.DIST.RT(-2,8)

由此，鮪魚罐頭的固形物重量的樣本平均數少於 138 公克的機率大約為 0.04，顯示鮪魚罐頭的固形物重量的樣本平均數有很高的可能性是不少於 138 公克。

例題 7.4

鮪魚罐頭製造商在罐頭標示固形物 140 公克。實際上，**固形物的重量是平均數 140 公克和標準差 4 公克**。假設你在賣場購買 36 罐的鮪魚罐頭，量測固形物的重量為樣本。試求：

(1) 樣本平均數的抽樣分配。

(2) 樣本平均數少於 138 公克的機率。

解 (1) 母體為鮪魚罐頭的固形物重量，平均數 140 公克和標準差 4 公克。

在賣場購買 36 罐的鮪魚罐頭為樣本，得罐頭的固形物重量 $X_1, \cdots, X_{36}$ 為獨立且有相同的機率分配，其平均數 140 公克和標準差 4 公克。

樣本平均數 $\bar{X} = \dfrac{X_1 + \cdots + X_{36}}{36}$ 的期望值與變異數分別為

$$E(\bar{X}) = 140$$

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{4^2}{36}$$

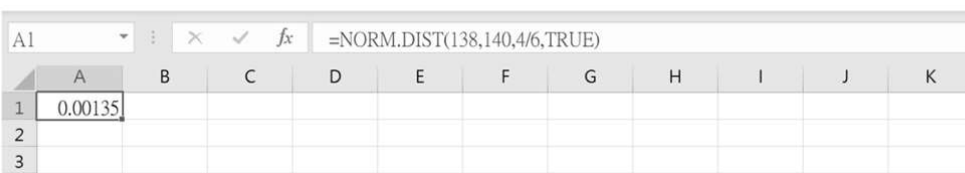
樣本數 $n = 36$，利用中央極限定理，可得

樣本平均數 $\bar{X}$ 的抽樣分配近似常態分配 $N(140\,,\,4^2/36)$

(2)

$$\frac{\bar{X} - 140}{4/\sqrt{36}} \xrightarrow{D} N(0\,,1)$$

$$P(\bar{X} < 135) = P\left(\frac{\bar{X}-140}{\frac{4}{\sqrt{36}}} < \frac{138-140}{\frac{4}{\sqrt{36}}}\right) \approx P(Z < -3) = 0.0013$$



Excel 2016 函式：=NORM.DIST(138,140,4/6,TRUE)

由此，鮪魚罐頭的固形物重量的樣本平均數少於 138 公克的機率大約為 0.001，顯示鮪魚罐頭的固形物重量的樣本平均數有很高的可能性是不少於 138 公克。

例題 7.5

鮪魚罐頭製造商在罐頭標示固形物 140 公克。實際上，**固形物的重量是平均數 140 公克和標準差 σ 未知**。假設你在賣場購買 36 罐的鮪魚罐頭，量測固形物的重量為樣本，得一組樣本標準差 s 為 3。試求：在賣場購買 36 罐的鮪魚罐頭固形物的樣本平均數少於 138 公克的機率。

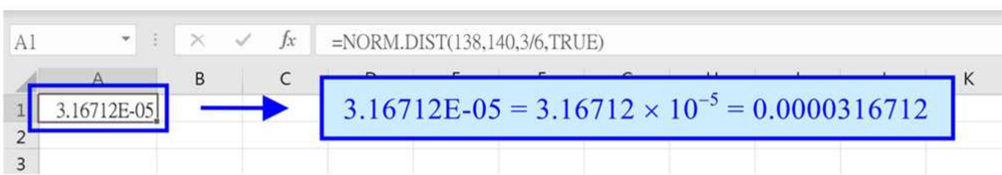
解 母體為鮪魚罐頭的固形物重量，平均數 140 公克和標準差 σ 未知。

在賣場購買 36 罐的鮪魚罐頭為樣本，得罐頭的固形物重量 $X_1, \cdots, X_{36}$ 為獨立且有相同的機率分配，其平均數 140 公克和標準差 σ 未知。

用樣本標準差 $s=3$ 估計母體標準差 σ，又 $n=36$ 為大樣本，所以

$$W = \frac{\bar{X}-140}{3/\sqrt{36}} \xrightarrow{D} N(0,1)$$

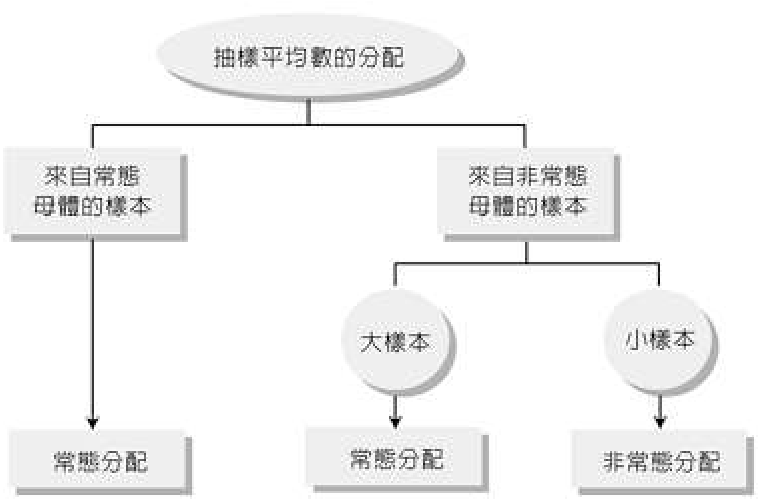
$$P(\bar{X} < 135) = P\left(\frac{\bar{X}-140}{\frac{3}{\sqrt{36}}} < \frac{138-140}{\frac{3}{\sqrt{36}}}\right) \approx P(W < -4) = 0.0000317$$



Excel 2016 函式：=NORM.DIST(138,140,3/6,TRUE)

由此，鮪魚罐頭的固形物重量的樣本平均數少於 138 公克的機率幾乎為 0，顯示鮪魚罐頭的固形物重量的樣本平均數有非常高的可能性是不少於 138 公克。

8-3-3 常態母體的樣本平均數抽樣分配



1. 若母體是常態分配，樣本平均數的抽樣分配是常態分配。

假設母體為常態分配$X \sim N(\mu, \sigma^2)$。 從此常態母體中，隨機且獨立的抽出大小為n的一組隨機樣本， 則隨機變數$X_1, X_2, ..., X_n$為獨立且有相同機率分配$N(\mu, \sigma^2)$。 由於常態分配具有相加特性，因此隨機變數樣本平均的機率分配仍是常態分配， 即$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$。

在計算樣本平均數$\bar{X}$抽樣分配的機率時，要將$\bar{X}$標準化成標準常態分配的隨機變數Z。

情況一：變異數σ^2已知

=> 當σ^2為已知時，樣本平均數的抽樣分配為常態分配，可透過標準化公式：$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$，計算機率。

倒題 7.6

設某學校學生身高呈常態分配，$\mu=160$，$\sigma=5$（ 公分 ），求：

(1) 隨機抽出 100 人，其平均身高 $\bar{X}$ 的抽樣分配。

(2) $P(\bar{X}<161)$、$P(159.5<\bar{X}<160.5)$ 及 $P(\bar{X}>161)$。

解：(1) $\bar{X}$ 為服從於 $N\left(\mu, \frac{\sigma^2}{100}\right)=N\left(160, \frac{25}{100}\right)$ 的常態分配，即 $\bar{X} \sim N\left(160, \frac{25}{100}\right)$

(2) 將 $\bar{X}$ 標準化成標準常態的隨機變數 Z，再討論各範圍的機率問題。

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{\bar{X}-160}{\sqrt{\frac{25}{100}}} \sim N(0,1)$$

(a) $P(\bar{X}<161)=P\left(Z<\frac{161-160}{\sqrt{\frac{25}{100}}}\right)=P(Z<2)=0.9772$

(b)
$$\begin{aligned} P(159.5<\bar{X}<160.5) &= P\left(\frac{159.5-160}{\sqrt{\frac{25}{100}}}<Z<\frac{160.5-160}{\sqrt{\frac{25}{100}}}\right) \\ &= P(-1<Z<1) \\ &= P(Z<1)-P(Z<-1) \\ &= 0.8413-0.1587=0.6826 \end{aligned}$$

(c) $P(\bar{X}>161)=P\left(Z>\frac{161-160}{\sqrt{\frac{25}{100}}}\right)=1-P(Z<2)$

$=1-0.9772=0.0228$

END

情況二：變異數σ^2未知

在計算樣本平均數$\bar{X}$抽樣分配的機率時，如果不知道母體變異數σ^2的值，就無法標準化為Z，透過標準常態分配來查表算機率。 此時，我們改用樣本變異數$S^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$， 利用t統計量，$t = \frac{\bar{X}-\mu}{S_{\bar{X}}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$，透過t分配查表算機率。 即$t \sim t_{n-1}$，其自由度為$n-1$。

=> 當σ^2為未知時，樣本平均數的抽樣分配為自由度為$n-1$的t分配。

(下面說明，可以省略。)

- 知，樣本平均數$\bar{X}$的抽樣分配為$N(\mu, \sigma^2/n)$，標準化後可得 $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- 又，統計量$Y = \frac{(n-1)S^2}{\sigma^2}$服從卡方分配，亦即$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
- 在統計學中，自由度（英語：degree of freedom, df）是指當以樣本的統計量來估計母體的母數時，樣本中獨立或能自由變化的數據的個數，稱為該統計量的自由度。一般來說，自由度等於獨立變數數減掉其衍生量數；舉例來說，變異數的定義是樣本減平均值（一個由樣本決定的衍生量）的平方之和，因此對n個隨機樣本而言，其自由度為n-1。
- 因為『若兩個隨機變數X與Y互為獨立，X的機率分配為$N(0,1)$，Y的機率分配為χ^2_k，則$\frac{X}{\sqrt{Y/k}}$的機率分配為t_k。』，所以$t = \frac{X}{\sqrt{\frac{Y}{n-1}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\frac{1}{n-1}}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$

例題 7.17

已知罐裝烏龍茶飲料呈常態分配，$\mu=15$ 元，但標準差未知，今自該分配抽出 $n=16$ 為一隨機樣本，樣本標準差 2 元，試求：

(1) $P(\bar{X}<14.1235)$

(2) $P(\bar{X}>16.301)$

(3) $P(14.1235<\bar{X}<16.301)$

解： $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ 為自由度 $16-1=15$ 的 t 分配。

利用 t 表，在自由度 15 的列中，找出指定的 t 值所對應的 α，即為我們要的 $P(t>t\text{ 值})$ 機率。

(1) $P(\bar{X}<14.1235)=P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}}<\frac{14.1235-15}{2/\sqrt{16}}\right)=P(t<-1.7530)=P(t>1.7530)$

$=0.05$

(2) $P(\bar{X}>16.301)=P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}}>\frac{16.301-15}{2/\sqrt{16}}\right)=P(t>2.602)=0.01$

(3) $P(14.1235<\bar{X}<16.301)=\left(\frac{14.1235-15}{2/\sqrt{16}}<\frac{\bar{X}-\mu}{S/\sqrt{n}}<\frac{16.301-15}{2/\sqrt{16}}\right)$

$=P(-1.7530<t<2.602)$

$=1-P(t>2.602)-P(t>1.7530)$

$=1-0.01-0.05=0.94$

END

=> 當σ^2為未知時，樣本平均數的抽樣分配為自由度為$n-1$的t分配，若為大樣本時，t分配會趨近於常態分配，請改用常態分配計算機率。

8-3-4 任意母體的樣本平均數抽樣分配

前提：大樣本

(1) 大數法則

當$n \to \infty$時，樣本平均數抽樣分配的變異數$Var(\bar{X}) = \frac{\sigma^2}{n} \to 0$，所以$\bar{X}$會趨近於$\mu$。 也就是說，隨著$n$越來越大，$\bar{X}$越來越沒有隨機性。在$n \to \infty$時，$\bar{X}$之抽樣分配會越來越集中，最後只剩下一個點$\mu$，此即大數法則的意義。

例題 7.9

設甲產品的重量標準差為 10 公克，今抽樣 100 件（採取出後放回抽樣），試求：

(1) 若 $n=100$ 時之標準誤。

(2) 若 $n=400$ 時之標準誤（樣本增大）。

(3) 若 $n=25$ 時之標準誤（樣本減小）。

(4) 第 (2)、(3) 小題的情況推論可靠性有何關係？

解：(1) $n=100$，$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$

(2) $n=400$，$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{400}} = \frac{1}{2}$

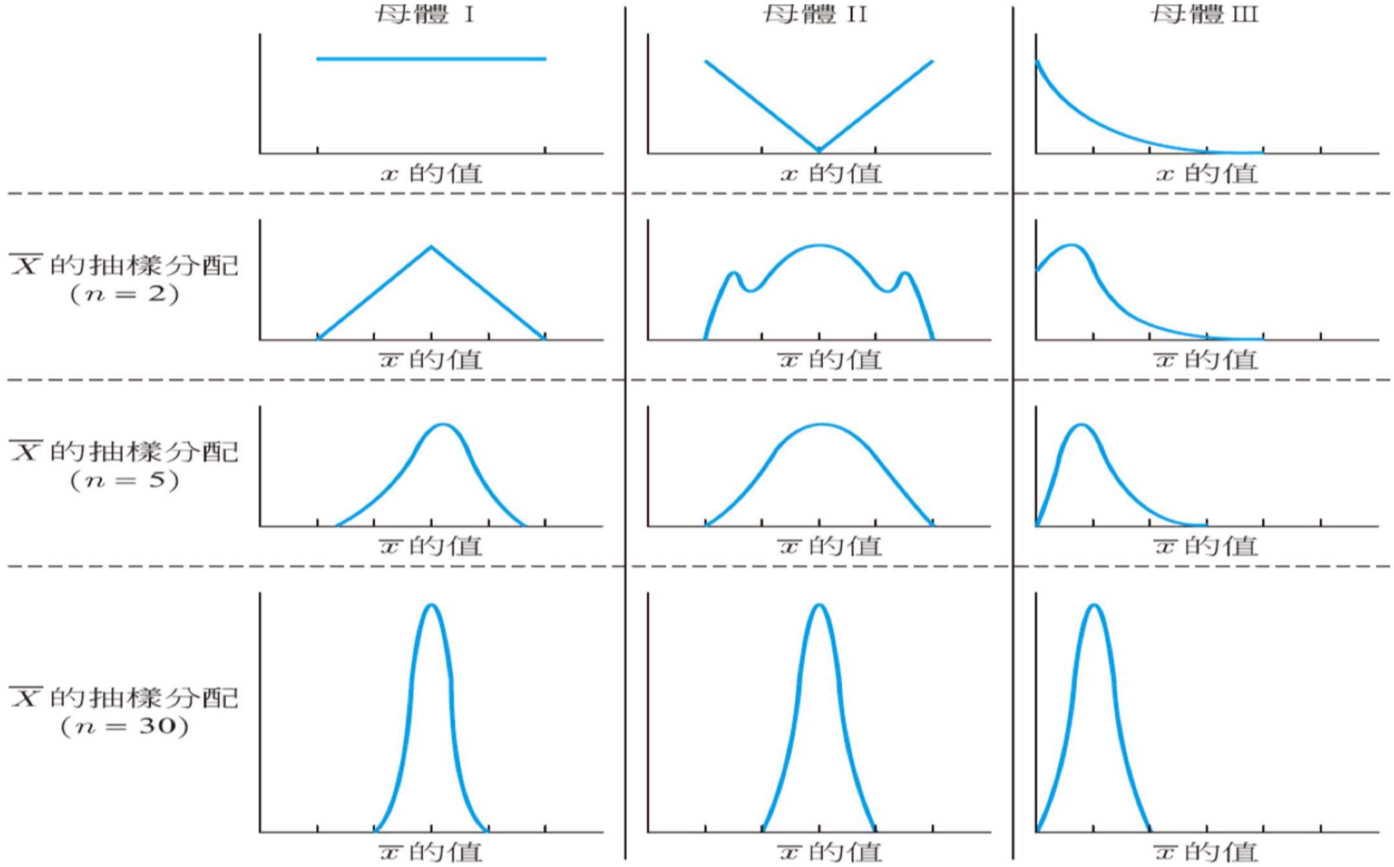
(3) $n=25$，$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$

(4) 由第 (2) 小題得知，當樣本數由 100 增大為 400 時，$\sigma_{\bar{X}}$ 由 1 減少為 1/2，標準誤減少，分散度減少，推論可靠性增加；由第 (3) 小題得知，樣本數由 100 減少為 25，$\sigma_{\bar{X}}$ 由 1 增加為 2，標準誤增加，分散度增加，推論可靠性減少。 END

(2)中央極限定理

當樣本數很大($n \geq 30$)時，不論母體之機率分配為何，$\bar{x}$的抽樣分配近似常態分配。

下面的左圖，例如是擲骰子，一次擲n個，$n = 2, 3, ...5, ... => n = 30$的變化



假設任意母體，其平均數為μ，變異數為σ^2。 從此母體隨機抽出大小為n的一組獨立樣本， 當樣本數n夠大時，樣本平均數$\bar{X}$的抽樣分配會近似常態分配 $N(\mu, \frac{\sigma^2}{n})$， 即$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$ 稱為中央極限定理。

=> 任意母體且變異數σ^2已知：

將$\bar{X}$標準化成標準常態分配的隨機變數Z，求算機率。 - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$

=> 任意母體且變異數σ^2未知：

用樣本變異數S^2替代母體變異數σ^2，求算機率。 - $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \rightarrow N(0,1)$ - t分配在樣本數夠大時，機率值趨近於常態分配，因此改用常態分配查表算機率。

例題 7.10

設從分配 $f(x)=\frac{1}{3}$，$x=1,2,3$ 中，以簡單隨機抽樣抽取 $n=100$ 的隨機樣本，試求 $\bar{X}$ 介於 1.5 到 2.5 之間的機率為何？

解：套用中央極限定理可知，$n=100$，$\bar{X}$ 的抽樣分配依然為常態分配，此常態分配有兩個參數，$E(\bar{X})=\mu$（原母體機率分配之期望值），$V(\bar{X})=\sigma^2/n$（原母體機率分配之變異數的 $1/n$ 倍）。因此我們須先計算原機率分配 $f(x)=\frac{1}{3}$，$x=1,2,3$ 之期望值與變異數。

$$E(X)=1\times\frac{1}{3}+2\times\frac{1}{3}+3\times\frac{1}{3}=2=\mu$$

$$V(X)=E(X^2)-[E(X)]^2=\left(1^2\times\frac{1}{3}+2^2\times\frac{1}{3}+3^2\times\frac{1}{3}\right)-2^2=\frac{2}{3}=\sigma^2$$

所以 $\bar{X}$ 抽樣分配之期望值與變異數分別為

$$E(\bar{X})=\mu=2$$

$$V(\bar{X})=\frac{\sigma^2}{n}=\frac{2/3}{100}=\frac{1}{150}$$

$\bar{X}$ 之抽樣分配為 $N\left(2,\frac{1}{150}\right)$。

欲求介於 1.5 到 2.5 之間的機率，需對 $\bar{X}$ 標準化

$$P(1.5\le\bar{X}\le 2.5)=P\left(\frac{1.5-2}{\sqrt{\frac{1}{150}}}\le Z\le\frac{2.5-2}{\sqrt{\frac{1}{150}}}\right)=P(-6.124\le Z\le 6.124)$$

$$=P(Z\le 6.124)-P(Z\le -6.124)\approx 1-0=1$$

END

例題 7.11

設有一個機率分配 $\mu=30$，$\sigma=5$，今自該分配抽出樣本數 $n=64$ 的隨機樣本，試求：

(1) $n=64$，$\bar{X}$ 抽樣分配之期望值與變異數。

(2) $P(\bar{X}<28.5)$。

(3) $P(\bar{X}>31)$。

(4) $P(29<\bar{X}<31.5)$。

解：套用中央極限定理，$\bar{X}$ 分配趨近常態分配

(1) 期望值：$E(\bar{X})=\mu=30$

變異數：$V(\bar{X})=\dfrac{\sigma^2}{n}=\dfrac{5^2}{64}=0.3906$

(2) $P(\bar{X}<28.5)=P\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\dfrac{28.5-30}{5/\sqrt{64}}\right)$

$=P(Z<-2.4)=0.0082$

(3) $P(\bar{X}>31)=P\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}>\dfrac{31-30}{5/\sqrt{64}}\right)$

$=P(Z>1.60)=1-P(Z<1.60)=1-0.9452=0.0548$

(4) $P(29<\bar{X}<31.5)=P\left(\dfrac{29-30}{5/\sqrt{64}}<\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\dfrac{31.5-30}{5/\sqrt{64}}\right)$

$=P(-1.60<Z<2.40)=P(Z<2.40)-P(Z<-1.60)$

$=0.9918-0.0548=0.9370$ END

~有些教科書在有限母體抽樣時，要計算平均數的變異數，會加上『有限母體修正因子』$\frac{N-n}{N-1}$。（$Var(\bar{X})=\frac{N-n}{N-1}\frac{\sigma^2}{n}$，$Var(\bar{p})=\frac{N-n}{N-1}\frac{p(1-p)}{n}$）。然而在許多條件下，有限母體可以視為無限母體處理，因而也可以省略處理。

8-4 樣本比例的抽樣分配

樣本比例$\bar{P}$亦是一種應用很廣的統計量，我們可以知道樣本比例的抽樣分配之後，就可以求特性，即樣本比例抽樣分配的期望值與標準誤，若樣本數夠大，也可以利用常態分配來算機率。

1. 抽樣分配的推導

假設母體為伯努力分配$B(1,p)$，其中p為母體中具有某種特性的比例。

從母體中，隨機且獨立的抽出大小為n的一組隨機樣本，則隨機變數$X_1,...,X_n$為獨立且有相同的機率分配$B(1,p)$，其期望值為$E(X_i)=p,$，變異數為 $V(X_i)=p(1-p),i=1,2,...,n$。

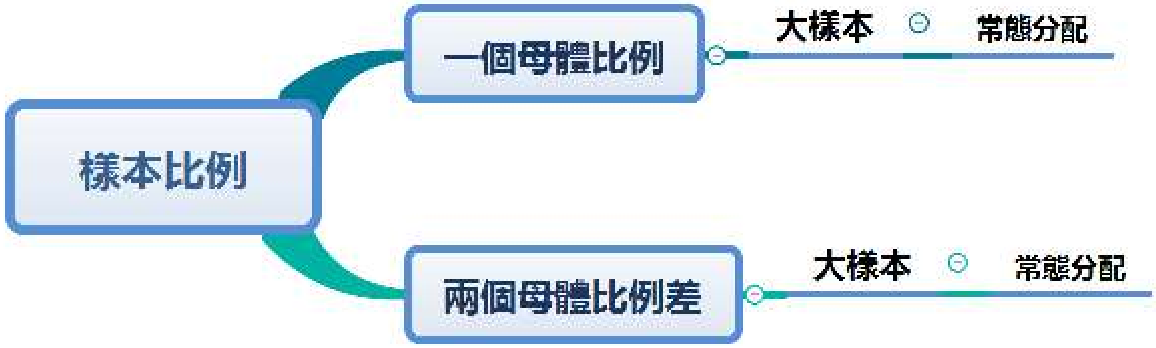
此時，樣本平均數$\bar{X}$為此組樣本具有該特性的樣本比例，即$\bar{X}=\frac{X1+...+Xn}{n}$，可用符號$\hat{p}$表示。

2. 期望值與變異數

樣本比例 $\hat{p}=\frac{X1+X2+...+Xn}{n}=\frac{\sum X}{n}=\bar{X}$，則樣本比例$\hat{p}$的期望值與變異數分別為

- 期望值：$E(\hat{p})=\frac{nE(X)}{n}=p$
- 變異數：$V(\hat{p})=\frac{nVar(X)}{n^2}=\frac{p(1-p)}{n}$

3. 樣本比例的抽樣分配趨近常態分配



利用中央極限定理，當樣本數n夠大，且$np > 5$和$n(1-p) > 5$時，可以得到樣本比例$\hat{p}$的抽樣分配會近似常態分配$N(p, p(1-p)/n)$。

因此，標準化隨機變數後，查標準常態分配表即可算機率。

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

=> 假設$X_1, ..., X_n$為獨立且來自相同的母體$B(1,p)$。當樣本數n夠大時，樣本比例$\hat{p} = \frac{X_1+...+X_n}{n}$的抽樣分配 會近似常態分配$N(p, p(1-p)/n)$，即$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to N(0,1)$。**

例題 7.7

某連鎖餐廳定期做顧客滿意度調查，根據這些調查，此連鎖店的經理對外宣稱有 80% 的顧客評定對整體服務為非常滿意。今隨機詢問 200 位顧客對該餐廳的服務滿意度調查，少於 75% 的顧客評定對整體服務為非常滿意的機率為何？

解 母體為此連鎖餐廳的顧客評定對整體服務為非常滿意的情況，服從伯努力分配 $B(1, 0.8)$。

隨機詢問 200 位顧客對該餐廳的服務滿意度調查，得樣本資料 200 筆，對整體服務為非常滿意的情況 $X_1, \cdots, X_{200}$ 為獨立且有相同的機率分配 $B(1, 0.8)$。

樣本比例 $\hat{p} = \frac{X_1 + \cdots + X_{200}}{200}$

$$E(\hat{p}) = 0.8$$

$$V(\hat{p}) = \frac{0.8(1-0.8)}{200} = 0.0008$$

樣本數 $n = 200$ 夠大，利用中央極限定理，得

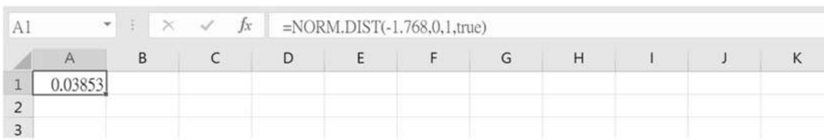
$$\hat{p} \xrightarrow{D} N(0.8\,,0.0008)$$

即

$$Z = \frac{\hat{p} - 0.8}{\sqrt{0.0008}} \xrightarrow{D} N(0\,,1)$$

少於 75% 的顧客評定對整體服務為非常滿意的機率：

$$P(\hat{p} < 0.75) = P\left(\frac{\hat{p} - 0.8}{\sqrt{0.0008}} < \frac{0.75 - 0.8}{\sqrt{0.0008}}\right) = P(Z < -1.768) = 0.03853$$



Excel 2016 函式：=NORM.DIST(-1.768,0,1,TRUE)

由此，評定對整體服務為非常滿意的結果中，「少於 75% 的顧客表示非常滿意」的機率大約為 0.04，顯示顧客表示「非常滿意」的比例有較高的可能性是高於 75%。