

# 樣本大小

## 3

Chapter



學習重點

- ☑ 樣本大小之選擇
- ☑ 估計平均數時的樣本大小
- ☑ 估計比率時的樣本大小

## 3-1 樣本大小之選擇

樣本不要過大，過大浪費成本；但也不要過小，過小則會有太大的抽樣誤差。如何決定適當的樣本大小？在機率抽樣的情況下，有關樣本大小的決定及樣本統計顯著性的判斷，可藉由機率法則的運用。（也就是說，有公式可供計算啦！）

但在非機率抽樣的情況下，除了依靠抽樣人員的主觀判斷或假設外，實無客觀之科學方法可資應用。

## 3-2 估計平均數時的樣本大小

### ▶▶ 母體變異數已知

於母體變異數（ $\sigma^2$ ）已知之情況下，樣本數（ $n$ ）之求算公式為：

$$n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{e} \right)^2$$

- $\alpha$  為顯著水準或風險水準， $(1-\alpha)$  即信賴係數或信賴水準
- $e$  為可容忍誤差
- $\sigma$  為母體標準差

先別忙著計算樣本數，由於本書是介紹 Excel 之書籍，故得對所使用到的各相關函數先介紹一下。

## ▶▶ 常態分配之 $z$ 值

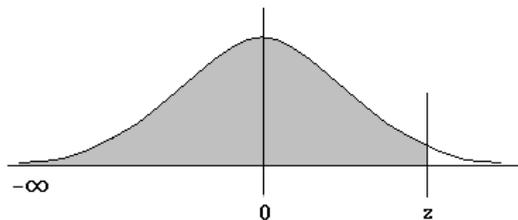
一般統計學之常態數值 ( $Z$ )，係利用查常態分配表 (附錄 A-2) 來得知。如： $Z_{0.025}$  為 1.96、 $Z_{0.05}$  為 1.645。但於 Excel 下，則可利用 NORM.S.INV() 標準常態分配反函數來查得；而若知道  $Z$  值，也可以 NORM.S.DIST() 函數來求得其機率。

## ▶▶ 標準常態分配 NORM.S.DIST()

標準常態分配函數之語法為：

```
NORM.S.DIST(z, 是否要累加)
NORM.S.DIST(z, cumulative)
```

cumulative 字面意思為累加，用以安排是否要累加？為 FALSE 時，其作用為求於標準常態分配 ( $\mu = 0, \sigma = 1$ ) 上，特定值  $z$  的機率。是否要累加為 TRUE 時，其作用為求自標準常態分配 ( $\mu = 0, \sigma = 1$ ) 的左尾開始，累加到  $z$  值處的總面積 (由  $-\infty$  積分到  $z$  後之結果)。即，下圖之陰影部份：



有了此函數，即可省去查常態分配表某  $z$  值之機率的麻煩。如：(詳範例 Ch03.xlsx 『NORM.S.DIST』工作表)

```
=NORM.S.DIST(-1.96, TRUE)    為 0.025
=NORM.S.DIST(-1.645, TRUE)   為 0.05
```

=NORM.S.DIST(0,TRUE) 為 0.5  
 =NORM.S.DIST(1.96,TRUE) 為 0.975

	A	B	C	D	E
1	常態分配				
2	z	自左尾累積			
3	-3.000	0.001			
4	-1.960	0.025			
5	-1.645	0.050	← =NORM.S.DIST(A5,TRUE)		
6	0.000	0.500	← =NORM.S.DIST(A6,TRUE)		
7	0.500	0.691			
8	1.000	0.841			
9	1.645	0.950			
10	1.960	0.975	← =NORM.S.DIST(A10,TRUE)		
11	3.000	0.999			



### 小秘訣

常態分配 (normal distribution) 是次數分配呈中間集中，而逐漸向左右兩端勻稱分散的鐘形曲線分佈。根據中央極限定理，不論原母體的分配為何？只要樣本數夠大 ( $n \geq 30$ )，樣本平均數  $\bar{X}$  的分配，會趨近於常態分配。

## ▶▶ 標準常態分配反函數 NORM.S.INV()

標準常態分配反函數之語法為：

NORM.S.INV(累計機率)  
 NORM.S.INV(probability)

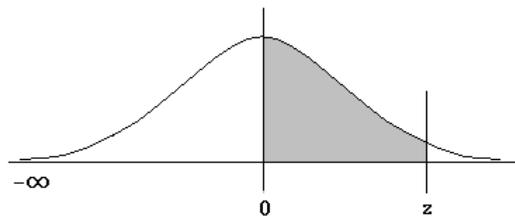
其作用為於標準常態分配 ( $\mu = 0, \sigma = 1$ )，求某累計機率所對應之 z 值。有了此函數，即可省去查常態分配表之 z 值的麻煩。如：(詳範例 Ch03.xlsx 『NORM.S.INV』工作表)

=NORM.S.INV(0.025) 為 -1.96  
 =NORM.S.INV(0.05) 為 -1.645  
 =NORM.S.INV(0.5) 為 0  
 =NORM.S.INV(0.95) 為 1.645  
 =NORM.S.INV(0.975) 為 1.96

	A	B	C	D	E
3	NORM.S.INV()				
4	自左尾累積	z			
5	0.001	-3.09			
6	0.025	-1.96	← =NORM.S.INV(A6)		
7	0.050	-1.64	← =NORM.S.INV(A7)		
8	0.100	-1.28			
9	0.250	-0.67			
10	0.500	0.00	← =NORM.S.INV(A10)		
11	0.600	0.25			
12	0.750	0.67			
13	0.900	1.28			
14	0.950	1.64	← =NORM.S.INV(A14)		
15	0.975	1.96	← =NORM.S.INV(A15)		
16	0.990	2.33			

## ►► 標準常態分配表

一般統計學之教科書，均會附有標準常態分配表（如：附錄 A-2），以利查常態數值（ $z$ ）。由於，常態分配是對稱的分配，故一般僅附上正值之部分，表內之累計機率，是由  $z$  值為 0 時開始累計。如： $z$  值 1.96，查得 1.96 之 0.475，表示由標準常態分配中央（ $z=0$ ）開始，累計到  $z=1.96$  的機率。即，下圖之陰影部份：



如要查負值之部份，仍以正值查表。然後，以 0.5 減去表內之累計機率即可。如： $Z$  值 -1.96，查得 1.96 之 0.475，以  $0.5-0.475=0.025$ ，即是自左尾開始累計到  $Z$  值為 -1.96 的機率。

相反地，若要計算由  $Z$  值為 -1.96 開始累計到右尾的機率，則將查得之值（0.475）加上 0.5，即 0.975。通常， $\alpha=0.05$  時，如要查  $Z_{\alpha/2}$  值，是找尋右尾機率為 0.025 時之  $Z$  值，即找出由左尾累積得 0.975 之  $Z$  值 1.96。若用 Excel 之 NORM.S.INV() 函數來求算，其公式應為：（詳範例 Ch03.xlsx 『依  $\alpha$  查  $Z$  值』工作表）

=NORM.S.INV(1-0.05/2)

	A	B	C	D	E
1	標準常態分配，均數為0，標準差為1				
2					
3	NORM.S.INV()				
4	$\alpha$ 值	$Z_{\alpha/2}$ 值			
5	0.01	2.576			
6	0.05	1.960			
7	0.10	1.645			
8	0.20	1.282			

於 Excel 下，利用 NORM.S.DIST() 函數即可輕易建立標準常態分配表。其建立步驟為：（詳範例 Ch03.xlsx 『常態分配表』工作表）

STEP 1 於 A2 輸入 Z 字串，A3 輸入 0.0（僅顯示 0），A4 輸入 0.1

	A	B	C
1		Z 值的小數第二位	
2	Z	0	0.01
3		0	
4		0.1	

STEP 2 選取 A3:A4，按『常用/數值/增加小數位數』 鈕，續按『常用/數值/減少小數位數』 鈕，使兩數均可擁有一位小數

	A
1	
2	Z
3	0.0
4	0.1

STEP 3 拖曳 A3:A4 右下角之複製控點，拉到 A33 位置，複製出 0.0、0.1、0.2、...、2.9、3.0 等數值

	A
30	2.7
31	2.8
32	2.9
33	3.0
34	

STEP 4 於 B1 輸入『Z 值的小數第二位』字串

STEP 5 於 B2 輸入 0.00（僅顯示 0），C2 輸入 0.01

- STEP 6 選取 B2:C2，按『常用/數值/增加小數位數』 鈕；續按『常用/數值/減少小數位數』 鈕，使兩數均可擁有 2 位小數

	A	B	C
1		Z 值的小數第二位	
2	Z	0.00	0.01
3	0.0		

- STEP 7 拖曳 B2:C2 右下角之複製控點，拉到 K2 位置，複製出 0.00、0.01、0.02、...、0.08、0.09 等數值

	H	I	J	K
1				
2	0.06	0.07	0.08	0.09

- STEP 8 於 B2:K2 尚呈選取之狀態，按『常用/儲存格/格式』 格式 鈕，續選「自動調整欄寬(I)」，將各欄調整成最適欄寬

	A	B	C	D	E	F	G	H	I	J	K
1		Z 值的小數第二位									
2	Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3	0.0										

- STEP 9 選取 B1:K1，按『常用/對齊方式/跨欄置中』 鈕，讓『Z 值的小數第二位』字串，於這幾欄內跨欄置中

	A	B	C	D	E	F	G	H	I	J	K
1		Z 值的小數第二位									
2	Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3	0.0										

- STEP 10 於 B3 輸入

```
=NORM.S.DIST($A3+B$2,TRUE)-0.5
```

	A	B	C	D	E	F	G	H	
1		Z 值的小數第二位							
2	Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	
3	0.0	0.00							

- STEP 11 拖曳其右下角之複製控點，往右複製到 K3

	A	B	C	D	E	F	G	H	I	J	K
1		Z 值的小數第二位									
2	Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3	0.0	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04

STEP 12 於 B3:K3 尚呈選取之狀態，按兩次『常用/數值/增加小數位數』 鈕，使各數均可有 4 位小數

STEP 13 於 B3:K3 尚呈選取之狀態，按『常用/儲存格/格式』 鈕之下拉鈕，續選「自動調整欄寬(I)」，調整成最適欄寬

	A	B	C	D	E	F	G	H	I	J	K
1		Z值的小數第二位									
2	Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3	0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
4	0.1										

STEP 14 雙按 K3 右下角之複製控點，將 B3:K3 往下複製到 K33，即完成整個建表工作

	A	B	C	D	E	F	G	H	I	J	K
		=NORM.S.DIST(\$A3+B\$2,TRUE)-0.5									
31	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
32	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
33	3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

## ▶▶ 母體變異數已知時的樣本大小

學過所需之幾個函數後，現在，可以

$$n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{e} \right)^2$$

來計算於母體變異數 ( $\sigma^2$ ) 已知之情況下的樣本數 (n)。式中：

- α 為顯著水準或風險水準，(1-α) 即信賴係數或信賴水準
- e 為可容忍誤差
- σ 為母體標準差

由這個公式中，可知其樣本大小決定於三個因素：

1. 母體變異數 ( $\sigma^2$ ) 的大小：母體變異數愈大，表其分散程度愈大，所需之樣本數就愈大。
2. 可容忍誤差 (e) 的大小：可容忍的誤差愈小，所需之樣本數就愈大。
3. 常態數值 (Z) 的大小：Z 值係由顯著水準 (α) 的大小來決定，α 愈小 Z 值愈大，Z 值愈大樣本數就愈大。

假定，母體變異數  $\sigma^2=6.25$  ( $\sigma=2.5$ )，於風險顯著水準  $\alpha=0.05$  ( $Z_{\alpha/2}=1.96$ ) 的情況下，希望對母體均數  $\mu$  的估計誤差  $e$  不超過 0.3，其樣本數應為多大？

將  $\sigma=2.5$  與  $Z_{\alpha/2}=1.96$  代入公式

$$n = \left( \frac{1.96 \times 2.5}{0.3} \right)^2 = 267$$

至少應取得 267 個樣本，才能有 95% 的保證其誤差不超過 0.3。

以 Excel 來安排相關之數字與公式，將為：（詳範例 Ch03.xlsx 『估計均數樣本數  $\sigma$  已知』工作表）

	A	B	C	D	E
1	母體變異數 $\sigma^2$	6.25			
2	母體標準差 $\sigma$	2.50	← =SQRT(B1)		
3	風險顯著水準 $\alpha$	0.05			
4	$Z_{\alpha/2}$	1.96	← =NORM.S.INV(1-B3/2)		
5	容忍誤差 $e$	0.30			
6	樣本數 $n$	267	← =((B4*B2)/B5)^2		

B 欄各儲存格之公式分別為：

母體標準差 $\sigma$	=SQRT(B1)
$Z_{\alpha/2}$	=NORM.S.INV(1-B3/2)
樣本數 $n$	=((B4*B2)/B5)^2

假定，電力公司根據過去之調查經驗，知道用戶用電度數的母體變異數為 48000 ( $\sigma^2=48000$ )、於顯著水準  $\alpha=0.05$  ( $Z_{\alpha/2}=1.96$ ) 的情況下，希望對母體平均用電度數  $\mu$  的估計誤差不超過 5 度，其樣本數應為多大？

將相關之數字輸入 Excel 即可算出，至少應取得 1844 個樣本，才能有 95% 的保證其估計誤差不超過 10 度：（詳範例 Ch03.xlsx 『估計均數樣本數  $\sigma$  已知 1』工作表）

	A	B	C	D	E
1	母體變異數 $\sigma^2$	48000			
2	母體標準差 $\sigma$	219	← =SQRT(B1)		
3	風險顯著水準 $\alpha$	0.05			
4	$Z_{\alpha/2}$	1.96	← =NORM.S.INV(1-B3/2)		
5	容忍誤差 $e$	10.00			
6	樣本數 $n$	1844	← =((B4*B2)/B5)^2		



以範例 Ch03.xlsx 『大學生零用金樣本數』工作表進行計算。假定，大學生之平均每月零用錢為 3,200 元，其變異數為 764,887。至少應取得多少樣本？才能有 95% 的信賴水準，保證其估計誤差不超過 100 元。

	A	B	C	D	E
1	母體變異數 $\sigma^2$	764887			
2	母體標準差 $\sigma$	875	← =SQRT(B1)		
3	風險顯著水準 $\alpha$	0.05			
4	$Z_{\alpha/2}$	1.96	← =NORM.S.INV(1-B3/2)		
5	容忍誤差e	100			
6	樣本數n	294	← =((B4*B2)/B5)^2		

### ▶▶ 平方根 SQRT() 函數

前例中，B2 之公式，使用到平方根 SQRT() 函數，其語法為：

SQRT(數值)  
SQRT(number)

本函數是用來求某數值的平方根，若數值為負值，本函數將回應 #NUM! 之錯誤。如：（詳範例 Ch03.xlsx 『平方根』工作表）

B2		: X ✓ fx =SQRT(A2)	
	A	B	C
1	X	X之平方根	
2	64	8	

事實上，有無此函數並不很重要。利用 ^ 運算符號也可達成開方之動作。如：=64^(1/2) 之結果即=SQRT(64)；但若要求開三方，那 SQRT() 可就無能為力了。但仍可利用 ^ 運算符號來解決（乘冪為 1/3 即等於開三方）：

C6		: X ✓ fx =A6^B6	
	A	B	C
5	X	次方	結果
6	64	1/3	4
7		↑	
8	此處是以0 1/3之方式所輸入之分數		

## ▶▶ 母體變異數未知的樣本大小

事實上，實務上很多數情況是無法得知母體變異數（ $\sigma^2$ ）。若是母體變異數未知，則可以過去調查之樣本變異數（ $S^2$ ）來替代。若過去也無類似之調查，可先做一小規模試訪，以利計算樣本變異數。然後，再來計算樣本數：

$$n = \left( \frac{Z_{\alpha/2} \cdot S}{e} \right)^2$$

式中：

- $\alpha$  為顯著水準或風險水準，（ $1-\alpha$ ）即信賴係數或信賴水準
- $e$  為可容忍誤差
- $S$  為樣本標準差

於 Excel 中，樣本變異數可以 VAR.S() 函數來求得，其語法為：

```
VAR.S(數值 1, 數值 2, ...)  
VAR.S(number1, [number2], ...)
```

number1,[number2],...為要計算變異數之儲存格或範圍引數，它是對應於某母體抽樣選出的 1 到 255 個數字引數樣本，方括號包圍之部份可省略。

樣本變異數（ $S^2$ ）的計算公式為：

$$\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}$$

其值恰為樣本標準差（ $S$ ）之平方，也是用來衡量觀測值與平均值間的離散程度。

由於，母體變異數未知。故舉行一次試訪，以範例 Ch03.xlsx 『估計均數樣本數  $\sigma$  未知』工作表，取得 120 位大學生之手機平均月費。計算出其樣本變異數（ $S^2$ ）為 314,592、於顯著水準  $\alpha=0.05$ （ $Z_{\alpha/2}=1.96$ ）的情況下，希望對母體手機平均月費  $\mu$  的估計誤差不超過 50 元，其樣本數應為多大？

將相關之數字輸入 Excel，即可算出，至少應取得 483 個樣本，才能有 95% 的保證其估計誤差不超過 50 元：

D6		fx		=((D4*D2)/D5)^2			
	A	B	C	D	E	F	G
1	平均月費		樣本變異數 $S^2$	314592	← =VAR.S(A2:A121)		
2	1610		樣本標準差 $S$	561	← =SQRT(D1)		
3	1080		風險顯著水準 $\alpha$	0.05			
4	1790		$Z_{\alpha/2}$	1.96	← =NORM.S.INV(1-D3/2)		
5	1520		容忍誤差 $e$	50			
6	480		樣本數 $n$	483	← =((D4*D2)/D5)^2		



由於，母體變異數未知。故舉行了一次試訪，取得 40 位大學生之信用卡每月平均簽帳金額。（詳範例 Ch03.xlsx 『信用卡問卷樣本數』工作表）於風險顯著水準  $\alpha=0.05$  ( $Z_{\alpha/2}=1.96$ ) 的情況下，希望對母體信用卡每月平均簽帳金額  $\mu$  的估計誤差不超過 100 元，其樣本數應為多大？

	A	B	C	D	E	F	G
1	每月平均簽帳金額					母體變異數 $\sigma^2$	834872
2	800	1800	600	500		母體標準差 $\sigma$	914
3	1200	2400	500	1200		風險顯著水準 $\alpha$	0.05
4	400	1600	2400	400		$Z_{\alpha/2}$	1.96
5	600	500	500	600		容忍誤差 $e$	100
6	2000	400	500	800		樣本數 $n$	321

### 3-3 估計比率時的樣本大小

若研究目的是在估計比率 ( $p$ , proportion)，其樣本數 ( $n$ ) 之求算公式為：

$$n = \left( \frac{Z_{\alpha/2} \cdot \sigma_p}{e} \right)^2$$

$p$  為母體的真正比率

$\alpha$  為風險顯著水準， $(1-\alpha)$  即信賴係數

$e$  為可容忍誤差

$\sigma_p$  為母體標準差，其運算公式為：

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

將其代入上式，即可獲致新的樣本數（n）公式：

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

不過，通常我們是無法得知母體之真正比率 p，要計算樣本數時，則以過去之調查結果替代。假定，上個月支持執政黨之比率為 38%(p)。這個月，於 95% 的信賴水準下（ $\alpha=0.05$ ），希望調查結果之支持率的允許誤差（e）為 3%，應取樣多少？

將相關值代入公式：

$$n = \frac{(1.96)^2 \times (0.38)(0.62)}{(0.03)^2} = 1006$$

至少應取得 1006 個樣本，才有 95% 的信心，保證其調查結果的支持率之誤差不超過  $\pm 3\%$ 。

以 Excel 來安排相關之數字與公式，將為：（詳範例 Ch03.xlsx 『以母體比率求樣本數』工作表）

B5		= (B3^2*B1*(1-B1))/B4^2			
	A	B	C	D	E
1	母體比率p	38%			
2	風險顯著水準 $\alpha$	0.05			
3	$Z_{\alpha/2}$	1.96	← =NORM.S.INV(1-B2/2)		
4	容忍誤差e	3%			
5	樣本數n	1006	← = (B3^2*B1*(1-B1))/B4^2		

## ►► 保守估計母體比率

於將前面之抽樣中，若將不同之  $p$  值分別代入，其樣本數勢必不同。茲將各種結果彙集成下表來比較：（詳範例 Ch03.xlsx 『以母體比率求樣本數』工作表）

	A	B	C	D	E	F
B8						$=(\$B\$3^2*A8*(1-A8))/\$B\$4^2$
1	母體比率p	38%				
2	風險顯著水準 $\alpha$	0.05				
3	$Z_{\alpha/2}$	1.96	$\leftarrow =\text{NORM.S.INV}(1-B2/2)$			
4	容忍誤差e	3%				
5	樣本數n	1006	$\leftarrow = (B3^2*B1*(1-B1))/B4^2$			
7	母體比率	樣本數				
8	30%	896	$\leftarrow = (\$B\$3^2*A8*(1-A8))/\$B\$4^2$			
9	38%	1006				
10	40%	1024				
11	45%	1056				
12	50%	1067				
13	55%	1056				
14	60%	1024				
15	70%	896				

可發現，其樣本數的極大值 1067 係發生於母體比率為 0.5 時。母體比率  $< 0.5$  時，隨母體比率逐漸增加，樣本數也逐步增加。母體比率  $> 0.5$  時，隨母體比率逐漸增加，樣本數則逐步減少。

故而，若我們無法得知母體真正比率  $p$ ，要計算樣本數時，可以採取最保守的估計，將母體比率設定為 0.5。這樣，由於其樣本數最大，所獲得之結果也將是各種情況下最正確的。



### 馬上練習

以範例 Ch03.xlsx 『以保守估計求樣本數』工作表進行計算，保守估計主張『開放核食』的支持度為 50%。要以 90% 之信賴水準（風險顯著水準  $\alpha=0.1$ ），希望調查結果之支持率的誤差為  $\pm 3\%$ ，應取樣多少？

	A	B
1	母體比率p	35%
2	風險顯著水準 $\alpha$	0.1
3	$Z_{\alpha/2}$	1.645
4	容忍誤差e	3%
5	樣本數n	684