# 第 11 章　統計估計

- **統計推論(statistical inference)**：利用樣本的訊息對母體做推論。
  - 例子：若想了解 2016 總統大選各可能候選人的支持度，最精確的方法當然是普查所有合格選民(普查沒有所謂的估計問題)，但普查會耗費過多成本。較佳的方法是，『從合格選民[母體]當中抽出一組樣本，運用該組樣本中所包含的訊息來估計各候選人的支持度，或是利用樣本資訊來檢定對候選人支持度的假設(如甲候選人的支持度是否過半)』。
    $\boxed{\text{問題}}$：(1)如何估計？(2)估計是否準確？[估計誤差如何？] (3)如何對某個假設進行檢定？
  - 推論統計包括兩大主軸：統計估計與假設檢定。
- **統計估計**(estimate)：利用樣本統計量去推估母體參數的方法。
- **假設檢定**(hypothesis testing)：對有關母體參數的假設，利用樣本的訊息，決定接受(不拒絕)該假設或拒絕該假設的統計方法。



## 點估計(Point estimate)

- 現實世界中需要運用統計估計的事件非常多，而所有估計問題的核心不外是：『我們想知道某些未知的母體參數之數值，但無法獲得母體的完整資料，故以一組樣本資料來推估這些母體參數』。
  - 某次大選中甲、乙兩候選人的支持度：母體參數為「所有合格選民中支持甲候選人的比例(母體比例)」，欲估計該母體參數，可從合格選民中抽出一組隨機樣本，以該樣本中兩候選人的支持率(樣本比例)來估計母體參數。
  - 台灣成年男性的平均身高：母體參數為「台灣所有成年男性身高的平均值(母體平均數)」，欲估計該母體參數，可從台灣成年男性中抽出一組隨機樣本測量其身高，以該樣本之身高平均值(樣本平均數)來估計母體參數。
- 在進行估計時，我們會從母體當中抽出一組隨機樣本，並根據該組樣本得出母體參數的估計。在呈現估計結果時有兩種方式：



  - 點估計：根據樣本計算出樣本統計量，以該樣本統計量做為母體參數的估計值。[點估計為一個實數值]
  - 區間估計：對未知的母體參數估計出一個上下限的區間，並指出該區間包含母體參數的可靠度。[區間估計為實數線上的一個區間]
- 點估計：由母體中抽取一組隨機樣本(樣本個數為 $n$)，並由該樣本中尋找一個樣本統計量來做為母體參數的估計值。
  - 正式說明：假設我們想估計某個母體未知參數 $\theta$，因此從母體中抽取一組個數為 $n$ 的隨機樣本 $X_1,\ldots,X_n$，並尋找樣本統計量 $\hat{\theta}$，若以該樣本統計量 $\hat{\theta}$ 來估計母體參數 $\theta$，則 $\hat{\theta}$ 即為 $\theta$ 的點估計。 [樣本統計量：$\hat{\theta}$ 為隨機樣本 $X_1,\ldots,X_n$ 的實數值函數，亦即 $\hat{\theta}=g(X_1,\ldots,X_n)$；其意義為，若知道隨機樣本 $X_1,\ldots,X_n$ 的確實觀察值，將其代入函數 $g$ 中，可得到一個實數值]



  - 名詞說明：樣本統計量這個名詞太廣義了；當我們在進行點估計時，用來估計母體參數的統計量通常稱為**估計式**(estimator)，將樣本觀察值代入估計式中所算出的實數值稱為**估計值**(estimate)。
  - 例子：以樣本平均數 $\bar{X}$ 來估計母體平均數 $\mu$，我們稱 $\bar{X}$ 為估計式；若我們已取得一組樣本，根據該樣本算出一個樣本平均數 $\bar{X}_0=125$，則 125 即為一個估計值。
- **點估計的步驟**

(1) 抽取具代表性的樣本
(2) 選擇一個較佳的樣本統計量做為估計式[統計理論大部分的工作都圍繞在這個問題上，因為統計估計必然會出現準確度的問題，我們當然希望一個估計式所提供的估計結果很準]
(3) 計算樣本統計量的值(估計值)
(4) 以樣本統計量的值推論母體參數值並做決策

- 例子：王太太想在台北市買一棟 25~30 坪的房子，因此想了解台北市的一般(平均)房價，該如何做呢？統計(點)估計
  - 第一步：從房屋仲介公司的 25~30 坪待售屋中抽取一組個數為 36 的隨機樣本，紀錄其價格(參閱 p.342 表 11.1)。
  - 第二步：以樣本平均數 $\bar{X}$ 來估計母體平均數 $\mu$，樣本標準差 $S$ 來估計母體標準差 $\sigma$。[這兩個估計式的準確度稍後再討論]
  - 第三步：以估計式所計算的樣本估計值為

    $$\bar{X} = \tfrac{1}{n}\sum_{i=1}^{n} X_i = 935.03 \quad (萬元)$$

    $$S = \sqrt{\tfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} = 223.25 \quad (萬元)$$

  - 第四步：根據上述樣本所計算的估計結果，王太太可得到底下結論：『台北市區 25~30 坪的房屋平均價格為 935.03 萬元，標準差為 223.25 萬元』。



## 點估計的限制

- 估計問題一定會引發一個疑慮：估計準確嗎？(估計誤差多大？)
  - 但是，到底要要如何判斷估計的準確度？
- 以點估計值來推估母體參數，無法判斷估計結果的真確性：
  - 假設我們知道台北市區 25~30 坪房屋的真實平均價格為 1000 萬元，則上例中的估計值 $\bar{X} = 935.03$ 很明顯是一個錯誤的估計值(估計誤差為 $\bar{X} - \mu = 1000 - 935.03 = -64.97$)；若抽取另一組樣本，該樣本所得到的估計值為 $\bar{X} = 1000$，則這個估計值「正中」母體參數。
  - 但我們根本無從得知母體參數真實數值，故根本無法判斷點估計值是否正中母體參數(亦無法得知估計誤差大小)。
  - 一般而言，點估計值總是與母體參數的真實值不同。
  - Remark：僅知道母體參數的點估計值無助於瞭解估計誤差的大小(亦即估計的準確度)，這也就是區間估計背後的動機



- 既然無法從估計值來判斷估計的準確度，那麼要如何判斷估計的好壞呢？————從**估計式**下手
  - 由於統計量(估計式)是隨機的，因此談論單一估計值有多準確，事實上是沒有意義的；應該衡量的是估計式的整體準確度，亦即站在還沒真正進行抽樣之前(事前)的角度來看估計式，衡量何種估計式(可視為是某特定估計方法)較準確。
  - 準確度越高的估計式就是越好的估計式：直覺上來講，一個好的估計式，其估計值接近真實參數的機率應該很高；而此時估計式的平均估計誤差、平均平方誤差(相對於其他估計式而言)應該較小。統計學正是從這個觀點來判斷估計式的好壞。[若一個估計式是好的，其估計的結果就可信]
  - 例子：若母體服從常態分配，則其平均數、中位數、眾數完全相同，但為何在估計母體平均數時我們總是用樣本平均數，而不是用樣本中位數、或樣本眾數；原因無他，樣本平均數是估計母體平均數的較好估計式(較準確、可信)。



## 估計式的評斷標準

- 符號：以 $\theta$ 表示(某個我們感興趣的)隨機變數之母體參數(是一個固定但未知的常數)，$\hat{\theta}$ 代表 $\theta$ 的估計式(隨機變數)。
- **估計誤差(estimation error)**：以 $\hat{\theta}$ 估計 $\theta$ 時，$\hat{\theta} - \theta$ 稱為估計誤差
- 判斷估計式優劣的直覺：良好估計式的估計誤差應該越小越好
  - 估計誤差有正有負，評估時應將估計誤差都變成正值(平方)，所有的可能的估計誤差均應納入考量(期望值)，這就導致了底下的評估準則。
  - 均方誤(mean squared error; MSE; 平均平方誤差)：一估計式 $\hat{\theta}$ 的均方誤定義為

    $$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

    口語上的解釋：誤差平方的平均值，可解釋為『估計式的平均誤差』。當然，MSE 越小代表估計式越準確。

- 均方誤可進一步拆解如下

$$
\begin{aligned}
MSE(\hat{\theta}) &= E[(\hat{\theta}-\theta)^2] = E[[(\hat{\theta}-E[\hat{\theta}])+(E[\hat{\theta}]-\theta)]^2] \\
&= E[(\hat{\theta}-E[\hat{\theta}])^2]+E[(E[\hat{\theta}]-\theta)^2]+E[2(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta)] \\
&= E[(\hat{\theta}-E[\hat{\theta}])^2]+E[(E[\hat{\theta}]-\theta)^2]+2(E[\hat{\theta}]-\theta)E[\hat{\theta}-E[\hat{\theta}]] \\
&= E[(\hat{\theta}-E[\hat{\theta}])^2]+E[(E[\hat{\theta}]-\theta)^2] \\
&= \underbrace{V(\hat{\theta})}_{\text{估計式的變異數}} + [\underbrace{E(\hat{\theta})-\theta}_{\text{估計式的偏誤}}]^2
\end{aligned}
$$

- MSE 由兩個非負值的部份組成：估計式的變異數 $V(\hat{\theta})$、估計式偏誤之平方 $[E(\hat{\theta})-\theta]^2$。因此，要使得 MSE 較小可從兩方面著手：『$V(\hat{\theta})$ 越小越好』、『$[E(\hat{\theta})-\theta]^2$ 越小越好』。
- 我們定義的第一個估計式評估準則『不偏性』，目的就在使得 $[E(\hat{\theta})-\theta]^2=0$。



**不偏性(unbiasedness)**

- **不偏估計式**(unbiased estimator)：若估計式 $\hat{\theta}$ 的平均數等於母體參數值 $\theta$ ［亦即 $E(\hat{\theta})=\theta$，或 $E(\hat{\theta}-\theta)=0$］，則稱該 $\hat{\theta}$ 為 $\theta$ 的不偏估計式，否則為**偏誤估計式**(biased estimator)。
  - 不偏性的意義：$\hat{\theta}-\theta$ 稱為**估計誤差**，不同的隨機樣本造成不同估計誤差，而這些估計誤差應有正有負；不偏性所代表的意義就是，若我們可進行無窮多次的隨機抽樣，則這無窮多次隨機抽樣所造成的估計誤差之平均值為 0；換句話說，這無窮多次估計的平均值 $E(\hat{\theta})$ 應等於真實參數 $\theta$；亦即，平均而言，估計式的實現值(估計值)會等於真實參數。
  - 若 $\hat{\theta}$ 為 $\theta$ 的不偏估計式，則 $\hat{\theta}$ 的 MSE 就是 $\hat{\theta}$ 的變異數，即

$$MSE(\hat{\theta})=V(\hat{\theta})+[Bias(\hat{\theta})]^2=V(\hat{\theta})$$

我們經常希望估計式是不偏的，以簡化估計式之 MSE 的計算，便於比較估計式之優劣，但並非所有估計式均具不偏性。

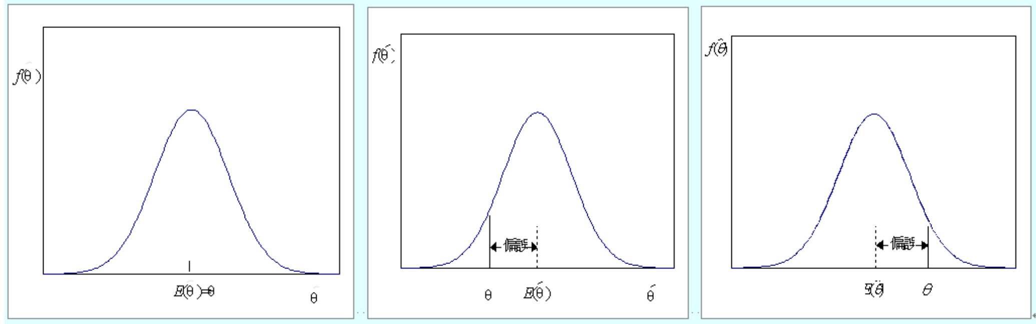


- **偏誤**(bias)：$E(\hat{\theta})$ 與 $\theta$ 的差距稱為偏誤，$Bias(\hat{\theta})=E(\hat{\theta})-\theta$

$Bias(\hat{\theta})=0 \Rightarrow$ 不偏(左圖)

$Bias(\hat{\theta})>0 \Rightarrow$ 正偏(中圖) $\Rightarrow$ 平均而言，估計值比真實參數大[高估參數值]

$Bias(\hat{\theta})<0 \Rightarrow$ 負偏(右圖) $\Rightarrow$ 平均而言，估計值比真實參數小[低估參數值]





- 例子：從平均數為 $\mu$、變異數為 $\sigma^2$ 的母體中隨機抽取 $n$ 個($n>4$)樣本 $X_1,\ldots,X_n$，考慮底下三個母體平均數 $\mu$ 的估計式之不偏性

$$T_1=\frac{X_1+2X_2+X_3}{4}$$

$$T_2=\frac{1}{4}X_1+\frac{X_2+\cdots+X_{n-1}}{2(n-2)}+\frac{1}{4}X_n$$

$$T_3=\frac{X_1+\cdots+X_n}{n}=\overline{X}$$

解答

$$E(T_1)=E\left[\frac{X_1+2X_2+X_3}{4}\right]=\frac{1}{4}E(X_1)+\frac{2}{4}E(X_2)+\frac{1}{4}E(X_3)$$

$$=\frac{1}{4}\mu+\frac{2}{4}\mu+\frac{1}{4}\mu=\mu \quad (不偏)$$

$$E(T_2)=E\left[\frac{1}{4}X_1+\frac{X_2+\cdots+X_{n-1}}{2(n-2)}+\frac{1}{4}X_n\right]$$

$$=\frac{1}{4}E(X_1)+\frac{E(X_2)+\cdots+E(X_{n-1})}{2(n-2)}+\frac{1}{4}E(X_n)$$

$$=\frac{1}{4}\mu+\frac{\overbrace{\mu+\cdots+\mu}^{(n-2)\text{項}}}{2(n-2)}+\frac{1}{4}\mu=\frac{1}{4}\mu+\frac{(n-2)\mu}{2(n-2)}+\frac{1}{4}\mu=\mu \qquad (\text{不偏})$$

$$E(T_3)=E\left[\frac{X_1+\cdots+X_n}{n}\right]=\frac{E(X_1)+\cdots+E(X_n)}{n}=\frac{\overbrace{\mu+\cdots\mu}^{n\text{項}}}{n}=\mu \quad (\text{不偏})$$

- ➢ 有可能一個參數的估計問題可以找到好幾個不偏估計式，所以僅根據不偏性無法判斷估計式的優劣[需考慮有效性]。
- ➢ 當然，也有可能某些參數無法找到不偏估計式，但卻可以找到具有「漸近不偏」以及「一致性」的估計式(後面說明)。



- 例子：從平均數為 $\mu$、變異數為 $\sigma^2$ 的母體中隨機抽取 $n$ 個($n>4$)樣本 $X_1,\ldots,X_n$ ，考慮底下兩個母體變異數 $\sigma^2$ 的估計式之不偏性

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

$$\hat{\sigma}^2=\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

**解答**

由於

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

$$=E\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right] \qquad \left\{\text{因}\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right\}$$

$$=\left[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\right]=\left[\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\tfrac{\sigma^2}{n}+\mu^2)\right]$$

$$\left\{\text{因}E(X_i^2)=V(X_i)+[E(X_i)]^2=\sigma^2+\mu^2;E(\bar{X}^2)=V(\bar{X})+[E(\bar{X})]^2=\tfrac{1}{n}\sigma^2+\mu^2\right\}$$

$$=n(\sigma^2+\mu^2)-(\sigma^2+n\mu^2)=(n-1)\sigma^2$$



因此

$$E(S^2)=E\left[\tfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\tfrac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

$$=\tfrac{1}{n-1}(n-1)\sigma^2=\sigma^2 \quad (\text{不偏})$$

$$E(\hat{\sigma}^2)=E\left[\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\tfrac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

$$=\tfrac{1}{n}(n-1)\sigma^2=\tfrac{n-1}{n}\sigma^2 \quad (\text{偏誤})$$

- 例子：樣本標準差 $S$ 並非母體標準差 $\sigma$ 的不偏估計式。

**解答**

因為 $V(S)=E(S^2)-[E(S)]^2=\sigma^2-[E(S)]^2$

$$\Rightarrow [E(S)]^2=\sigma^2-V(S)$$

由於 $V(S)>0$ [因為是 $S$ 的變異數]，因此

$$[E(S)]^2=\sigma^2-V(S)<\sigma^2$$

故 $E(S)<\sigma$ ，具有偏誤。



**有效性(efficiency)**

- 均方誤(MSE)所衡量的是估計式的平均誤差大小，因此 MSE 才是衡量估計式優劣的良好準則；當然，一個估計式的 MSE 較小(平均誤差較小)，就代表是較好的估計式。
- **絕對有效性**(absolute efficiency)：假設 $\hat{\theta}$ 為 $\theta$ 的估計式，若 $\hat{\theta}$ 的均方誤 $MSE(\hat{\theta})=E[(\hat{\theta}-\theta)^2]$ 為所有的估計式中最小者，則稱 $\hat{\theta}$ 在估計 $\theta$ 時具有絕對有效性。
  - ➢ 要找出具有絕對有效性的估計式非常困難，因為必須把所有可能估計式找出來，比較其 MSE 後方可找到具絕對有效性的估計式。因此，在探討有效性時，通常只會討論少數幾個估計式的相對有效性(如侷限在不偏估計式)。
- **相對有效性**(relative efficiency)：假設 $\hat{\theta}$、$\hat{\hat{\theta}}$ 均為 $\theta$ 的估計式，若 $\hat{\theta}$ 的均方誤相對 $\hat{\hat{\theta}}$ 的均方誤較小，亦即：

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

則 $\hat{\theta}$ 相對 $\hat{\hat{\theta}}$ 為有效估計式。

➢ 如果估計式是不偏的，則估計式的 MSE 就等於估計式的變異數，亦即 $MSE(\hat{\theta}) = V(\hat{\theta})$。比較不偏估計式的 MSE 相對簡單，因此相對有效性通常應用於不偏估計式的優劣比較。

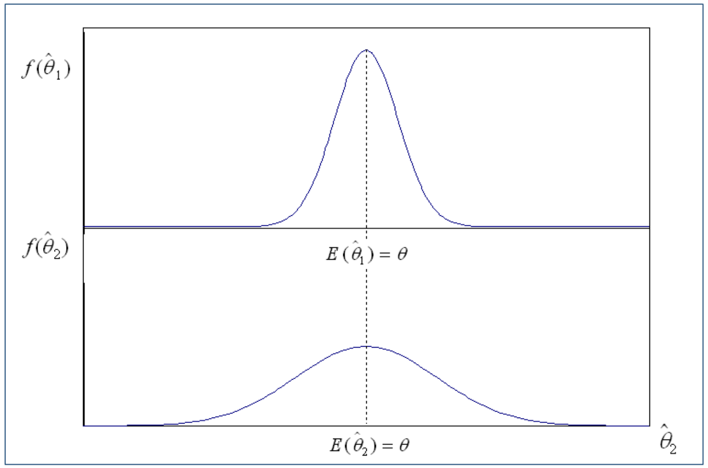
- 不偏估計式的相對有效性：假設 $\hat{\theta}$、$\hat{\hat{\theta}}$ 均為參數 $\theta$ 的不偏估計式，若

$$\frac{V(\hat{\theta})}{V(\hat{\hat{\theta}})} < 1$$

則 $\hat{\theta}$ 相對 $\hat{\hat{\theta}}$ 為有效估計式。



- 如下圖中，$\hat{\theta}_1$ 與 $\hat{\theta}_2$ 均為 $\theta$ 的不偏估計式，但 $\hat{\theta}_1$ 的分散程度(變異數)比 $\hat{\theta}_2$ 小[這不就是代表 $\hat{\theta}_1$ 靠近真實參數 $\theta$ 的機率比 $\hat{\theta}_2$ 高嗎？]，因此在估計 $\theta$ 時，$\hat{\theta}_1$ 相對於 $\hat{\theta}_2$ 具有效性。






- 例子：考慮底下兩母體平均之估計式的相對有效性

$$T_1 = \frac{X_1 + 2X_2 + X_3}{4}$$

$$T_3 = \frac{X_1 + \cdots + X_n}{n} = \bar{X}$$

**解答** 由於 $T_1$ 與 $T_3$ 均為不偏估計式，因此其 MSE 即為其變異數

$$MSE(T_1) = V(T_1) = V\left[\frac{X_1 + 2X_2 + X_3}{4}\right]$$

$$= \tfrac{1}{16}[V(X_1) + 4V(X_2) + V(X_3)] = \tfrac{1}{16}(\sigma^2 + 4\sigma^2 + \sigma^2) = \tfrac{6}{16}\sigma^2 = \tfrac{3}{8}\sigma^2$$

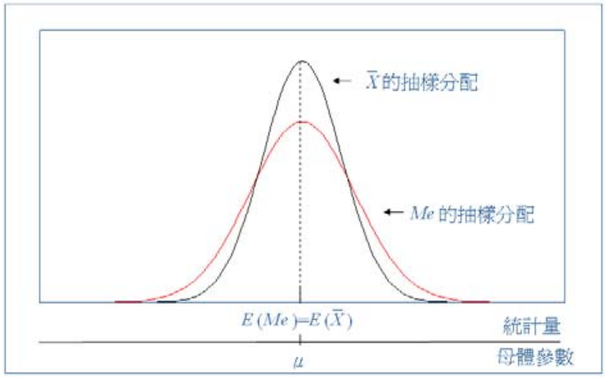
$$MSE(T_3) = V(T_3) = V(\bar{X}) = \tfrac{1}{n}\sigma^2$$

故 $\dfrac{MSE(T_3)}{MSE(T_1)} = \dfrac{\frac{1}{n}\sigma^2}{\frac{3}{8}\sigma^2} = \frac{8}{3n} < 1 \quad$ (若 $n > \frac{8}{3}$)

由於之前已假設 $n > 4$，故「$T_3$ 相對 $T_1$ 在估計 $\mu$ 時具相對有效性」



- 例子：若母體具常態分配，樣本平均數 $\bar{X}$ 與樣本中位數 $m_e$ 均是母體平均數的不偏估計式；但兩者的變異數分別為 $V(\bar{X}) = \frac{\sigma^2}{n}$、$V(m_e) = \frac{\pi}{2}\frac{\sigma^2}{n}$；因此 $\dfrac{V(\bar{X})}{V(m_e)} = \dfrac{2}{\pi} < 1$，故 $\bar{X}$ 相對 $m_e$ 在估計母體平均數時具相對有效性(這就是為什麼選擇用 $\bar{X}$ 來估計 $\mu$ 的原因)



- 例子：考慮底下兩個常態母體變異數估計式的相對有效性

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

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

**解答** 我們已知 $E(S^2)=\sigma^2$、$E(\hat{\sigma}^2)=\frac{n-1}{n}\sigma^2$，因此 $Bias(S^2)=0$、$Bias(\hat{\sigma}^2)=\frac{-1}{n}\sigma^2$；而且我們可以證明 $V(S^2)=\frac{2\sigma^4}{n-1}$。又由於

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{n-1}{n}S^2$$

因此 $V(\hat{\sigma}^2)=V(\frac{n-1}{n}S^2)=\frac{(n-1)^2}{n^2}V(S^2)=\frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1}=\frac{(n-1)2\sigma^4}{n^2}$

故 $MSE(S^2)=V(S^2)+[Bias(S^2)]^2=\frac{2\sigma^4}{n-1}$

$MSE(\hat{\sigma}^2)=V(\hat{\sigma}^2)+[Bias(\hat{\sigma}^2)]^2=\frac{(n-1)2\sigma^4}{n^2}+\frac{\sigma^4}{n^2}=\frac{(2n-1)\sigma^4}{n^2}$

所以 $\dfrac{MSE(\hat{\sigma}^2)}{MSE(S^2)}=\dfrac{\frac{(2n-1)\sigma^4}{n^2}}{\frac{2\sigma^4}{n-1}}=\dfrac{(2n-1)(n-1)}{2n^2}=(1-\frac{1}{2n})(1-\frac{1}{n})<1$

故在估計 $\sigma^2$ 時，$\hat{\sigma}^2$ 相對於 $S^2$ 有效。



**最小變異不偏性(minimum variance unbiasedness: MVU)**

- 雖然絕對有效估計式不容易找到，但在不偏估計式的集合當中我們卻可以較容易找到一個變異數最小的估計式。
- **最小變異不偏估計式(minimum variance unbiased estimator; MVUE)**：若一個母體參數 $\theta$ 的不偏估計式 $\hat{\theta}$，其變異數是所有 $\theta$ 之不偏估計式中最小者，則稱 $\hat{\theta}$ 為最小變異不偏估計式(MVUE)。
  - MVUE 的正式定義：$\hat{\theta}$ 為 $\theta$ 之 MVUE，若且為若

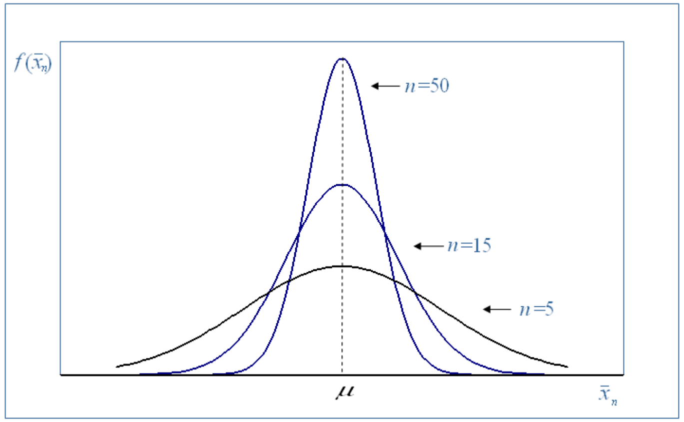
    (1) $E(\hat{\theta})=\theta, \quad \forall\theta$。

    (2) 若 $\hat{\theta}^*$ 亦為不偏估計式，則 $V(\hat{\theta})\le V(\hat{\theta}^*)$。
- 不偏性、有效性、最小變異不偏性均是小樣本性質，不管樣本個數 $n$ 多大均成立的性質。但有某些估計式在小樣本時不具上述三個性質，但只要樣本個數 $n$ 增加($n\to\infty$)時這些估計式即具有良好的大樣本性質，我們依舊將這些估計式視為優良估計式。



**一致性(consistency)**

- **一致性估計式**(consistent estimator)：設 $\hat{\theta}_n$ 為 $\theta$ 的估計式(下標 $n$ 代表樣本觀察值個數)，若 $\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|\le\varepsilon)=1$，則 $\hat{\theta}_n$ 為 $\theta$ 之一致性估計式($\varepsilon$ 為正的極小數值)。






- 意義：當我們用來估計參數的樣本觀察值個數 $n$ 增大時，估計直非常靠近真實參數的機率為 1。
- 直覺的解釋：例如當我們在估計母體平均數時(母體個數無限)，若取得的樣本個數 $n$ 越來越多，樣本的特性應該會跟母體的特性越來越近似，樣本平均數也會非常接近母體平均數。[由於 $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，當樣本個數 $n\to\infty$ 時 $\frac{\sigma^2}{n}\to 0$，變異數為 0 的隨機變數就是固定常數，該常數就是 $\mu$]
- 這就表示增加樣本觀察值個數 $n$ 可提高估計的準確程度；一致性是一個很直覺的特性，因此通常我們都會要求估計式要具有一致性。
- $\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|\le\varepsilon)=1$ 可簡化表達為

  $\text{plim}\,\hat{\theta}_n=\theta$ 或 $\hat{\theta}_n\xrightarrow{P}\theta$

  我們會說『$\hat{\theta}_n$ 在機率上收斂(converges in probability to)到 $\theta$』，$\theta$ 稱為 $\hat{\theta}_n$ 的機率極限(probability limit)。

- 要根據上述定義驗證一個估計式是否為一致性估計式很麻煩，底下介紹一個比較簡單的方法，大部分估計式的一致性都可透過這個方法來加以驗證。
- 驗證一致性估計式的定理：
  - **MSE 一致性**(MSE consistency)：若

    $$\lim_{n\to\infty} MSE(\hat{\theta}_n) = \lim_{n\to\infty} E[(\hat{\theta}_n - \theta)^2] = 0$$

    我們稱 $\hat{\theta}_n$ 為 $\theta$ 的 **MSE 一致估計式**，通常表達為

    $$\hat{\theta}_n \xrightarrow{MSE} \theta$$

  - **漸近不偏**(asymptotic unbiasedness)：若

    $$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

    則稱 $\hat{\theta}_n$ 為漸近不偏。
    - 特例：若 $\hat{\theta}$ 為不偏估計式，則 $\hat{\theta}$ 亦為漸近不偏。



- MSE 一致性的充要條件：
  $\hat{\theta}_n \xrightarrow{MSE} \theta$ 若且為若(if and only if)

  $$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta \qquad 且 \qquad \lim_{n\to\infty} V(\hat{\theta}_n) = 0$$

- 一致性的充分條件：若 $\hat{\theta}_n \xrightarrow{MSE} \theta$，則 $\hat{\theta}_n \xrightarrow{P} \theta$；亦即估計式為一致性的充分條件為『MSE 一致性』。
  - 簡單而言，只要一個估計式滿足底下兩個條件，就是一致性估計式

    $$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta \qquad 以及 \qquad \lim_{n\to\infty} V(\hat{\theta}_n) = 0$$

- 另一個跟一致性有關的有用定理(**Slutsky 定理**)：
  - 若 $\hat{\theta}_n$ 為 $\theta$ 的一致性估計式，則對於任何的連續函數 $g(\cdot)$ 而言，$g(\hat{\theta}_n)$ 亦為 $g(\theta)$ 的一致性估計式。
  - 推展：若 $\hat{\theta}_{1n}$ 與 $\hat{\theta}_{2n}$ 分別為 $\theta_1$ 與 $\theta_2$ 的一致性估計式，則對於任何的雙變數連續函數 $g(\cdot,\cdot)$ 而言，$g(\hat{\theta}_{1n}, \hat{\theta}_{2n})$ 亦為 $g(\theta_1, \theta_2)$ 的一致性估計式。



- 例子：樣本平均數 $\bar{X}$ 為母體平均數 $\mu$ 的一致性估計式

  **解答** 由於 $E(\bar{X}) = \mu$ 、$V(\bar{X}) = \frac{\sigma^2}{n}$ ，因此

  $$\lim_{n\to\infty} E(\bar{X}) = \lim_{n\to\infty} \mu = \mu$$

  $$\lim_{n\to\infty} V(\bar{X}) = \lim_{n\to\infty} \tfrac{\sigma^2}{n} = 0$$

  因此樣本平均數 $\bar{X}$ 為母體平均數 $\mu$ 的一致性估計式。

- 例子：樣本變異數 $S^2$ 為母體變異數 $\sigma^2$ 的一致性估計式，樣本標準差 $S$ 為母體標準差 $\sigma$ 的一致性估計式。

  **解答** 由於 $E(S^2) = \sigma^2$ 、$V(S^2) = \frac{2\sigma^4}{n-1}$ ，因此

  $$\lim_{n\to\infty} E(S^2) = \lim_{n\to\infty} \sigma^2 = \sigma^2$$

  $$\lim_{n\to\infty} V(S^2) = \lim_{n\to\infty} \tfrac{2\sigma^4}{n-1} = 0$$

  故樣本變異數 $S^2$ 為母體變異數 $\sigma^2$ 的一致性估計式。



由於 $g(x) = \sqrt{x}$ 為 $x$ 的連續函數，根據 Slutsky 定理，$\sqrt{S^2}$ 為 $\sqrt{\sigma^2}$ 的一致性估計式，亦即，樣本標準差 $S$ 為母體標準差 $\sigma$ 的一致性估計式[然而樣本標準差 $S$ 卻不是母體標準差 $\sigma$ 的不偏估計式]。

- 例：樣本變異係數 $\frac{S}{\bar{X}}$ 為母體變異係數 $\frac{\sigma}{\mu}$ 的一致性估計式($\mu \neq 0$)
  - 由於 $\bar{X}$ 為 $\mu$ 的一致性估計式，$S$ 為 $\sigma$ 的一致性估計式，且 $g(x,y) = \frac{x}{y}$ 為 $x$ 與 $y$ 的連續函數，根據 Slutsky 定理，$\frac{S}{\bar{X}}$ 為 $\frac{\sigma}{\mu}$ 的一致性估計式。

**有關估計式評估的總結**

- 若估計式具有良好的小樣本性質(不偏性、有效性)與大樣本性質(一致性)，當然就是最好的估計式。但有時……事與願違！！某些估計式不具不偏性，這時我們會要求它至少須具有一致性。
- 有沒有一些好的『一般性估計方法』，可確保我們對參數的估計具有良好的小樣本與大樣本統計性質呢？當然有，底下開始介紹

## 參數估計方法

**最小平方法(least squares method; LS)[普通最小平方法(ordinary least squares); OLS]**

- LS 的估計原理：找到一個估計式，使得估計值與觀察值的誤差平方和極小化。
  - 將誤差取平方的理由有二：
    - (1) 在加總誤差時避免正負號相抵消。
    - (2) 取平方等於是在估計時重視較大的誤差(因絕對數值越大，平方越大)。
- 以最小平方法求得的估計式稱為最小平方估計式(least squares estimator; LSE)。LSE 具有底下兩個特性：
  - LSE 具有不偏性、一致性，且為最佳線性不偏估計式(best linear unbiased estimator; BLUE)。BLUE 是指『在所有的線性不偏估計式中，變異數(也指 MSE)最小的估計式』。
  - 最小平方法在估計時不需知道母體的機率分配。



- 例子：假設有一平均數為 $\mu$ 的母體，$X_1,...,X_n$ 為來自於該母體的一組隨機樣本，則母體平均數 $\mu$ 的最小平方估計式為何？
  - LSE 必須使得觀察值($X_i$)與估計式之值($\hat{\mu}$)的誤差平方和最小，亦即必須求解底下問題

    $$Min \quad Q=\sum_{i=1}^{n}(X_i-\hat{\mu})^2$$

    將 $Q$ 對 $\hat{\mu}$ 取一階微分並令其等於 0 可得[一階條件]

    $$\frac{\partial Q}{\partial \hat{\mu}}=-2\sum_{i=1}^{n}(X_i-\hat{\mu})=0$$

    因此 $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$ [樣本平均數其實是 LSE]
  - 在母體平均數與變異數的估計問題中，LS 僅可找出母體平均數的估計式，至於母體變異數通常直接設定一個不偏估計式，例如樣本變異數：

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



**動差法(method of moments; MM)**

- **母體 $r$ 級原始動差**：設 $X$ 為一隨機變數，其機率(密度)函數為 $f(x;\theta)$，其中 $\theta$ 為母體參數，則 $X$ 的母體 $r$ 級原始動差為

  $$U_r=E(X^r),\ r=1,2,3\cdots$$

- **樣本 $r$ 級原始動差**：設 $X_1,...,X_n$ 為一組樣本，則該樣本的樣本 $r$ 級原始動差為

  $$U'_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r,\ r=1,2,3\cdots$$

- MM 的估計原理：以樣本動差來取代母體動差，藉此找出母體參數的估計式。以動差法所求出的估計式稱為**動差估計式**(MME)。
  - 若母體 $X$ 具有機率(密度)函數 $f(x;\theta)$，其中 $\theta$ 為該分配之參數，則母體 $X$ 的動差一定是母體參數 $\theta$ 的函數。
  - MM 即透過 $U'_1=U_1$、$U'_2=U_2$、…等動差條件來找出母體參數的估計式[當然，須估計幾個參數就要有幾個動差條件]



- 例子：假設有一母體 $X\sim N(\mu,\sigma^2)$，$X_1,...,X_n$ 為自 $X$ 中抽出的一組隨機樣本，則 $\mu$ 與 $\sigma^2$ 的動差估計式為何？
  - 由於需要估計兩個參數($\mu$ 與 $\sigma^2$)，因此需要兩個動差條件；我們使用一級與二級動差。
  - 母體動差

    $$U_1=E(X)=\mu$$

    $$U_2=E(X^2)=V(X)+[E(X)]^2=\sigma^2+\mu^2$$
  - 動差條件：樣本動差＝母體動差

    $$U'_1=U_1\Rightarrow\frac{1}{n}\sum_{i=1}^{n}X_i=\mu$$

    $$U'_2=U_2\Rightarrow\frac{1}{n}\sum_{i=1}^{n}X_i^2=\sigma^2+\mu^2$$
  - 由第一個方程式可解得[以 $\hat{\mu}_{MME}$ 代表 $\mu$ 的 MME]

    $$\hat{\mu}_{MME}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}=U'_1 \text{ [樣本動差的函數]}$$

- 第二個方程式可解得[以 $\hat{\sigma}^2_{MME}$ 代表 $\sigma^2$ 的 MME]

$$\hat{\sigma}^2_{MME}=U_2'-(U_2')^2=\tfrac{1}{n}\sum_{i=1}^{n}X_i^2-(\tfrac{1}{n}\sum_{i=1}^{n}X_i)^2$$
$$=\tfrac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=\tfrac{1}{n}\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right]=\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

[這也是樣本動差的函數；該估計式與樣本變異數不同]

- 疑問：為什麼用『一級與二級動差』？為什麼不用其他的動差條件，如『一級與四級動差』？當然可以，只不過估計式會不相同，估計值亦會不同[這是 MME 另一個大問題]

● 例子：假設母體 $X$ 服從指數分配，分配之參數為 $\lambda$；自 $X$ 中隨機抽取一組樣本 $X_1,...,X_n$，試求 $\lambda$ 的動差估計式。

- 由於僅需估計一個參數($\lambda$)，故利用一級動差條件來估計
- 指數分配的一級母體動差為[指數分配 pdf 為 $f(x)=\lambda e^{-\lambda x}$]

$$U_1=E(X)=\tfrac{1}{\lambda}$$



- 動差條件：$U_1'=U_1 \Rightarrow \tfrac{1}{n}\sum_{i=1}^{n}X_i=\tfrac{1}{\lambda}$
- 因此可解得：$\hat{\lambda}_{MME}=\dfrac{1}{\tfrac{1}{n}\sum_{i=1}^{n}X_i}=\dfrac{1}{\bar{X}}$

● 動差估計式的性質：

- 使用動差法時必須假設母體的分配函數 $f(x;\theta)$ 已知[如此才能知道母體動差與什麼參數有關]。
- 動差估計式為樣本動差的函數，又由於樣本動差為一致性估計式[一般化的大數法則：$\tfrac{1}{n}\sum_{i=1}^{n}X_i^r \xrightarrow{p} E(X^r)$]，因此根據 Slutsky 定理，動差估計式為母體參數的一致性估計式[但 MME 不具不偏性]。
- MME 的抽樣分配通常具漸近常態的性質(CLT 的應用)。
- MME 通常不具漸近有效性[即使樣本個數 $n\to\infty$，MSE 還是相對比其他估計式高]；因此 Hansen (1982)提出一般化動差法(Generalized Method of Moment; GMM)來改進此缺點。



**最大概似法(maximum likelihood method; ML)**

● 估計原理：所尋找的估計式，目的在使得出現該筆樣本觀察值的機率(概似值)是最高的。

● 暖身的例子：假設盒中有 5 個球，但不知有幾個紅球幾個白球。現要估母體參數 $\theta$ (紅球個數)。自盒中抽取 3 個球為樣本(樣本數 $n=3$ 的樣本)，若抽出的 3 個球均為紅色，則該樣本(3 個紅球)出現的機率，取決於母體參數 $\theta$ (盒中紅球數)。下表列出各種母體參數 $\theta$ 下出現『3 個紅球』的機率。

- ML 所找出的估計值為『出現 3 個紅球的機率是最高的』參數值，亦即 $\hat{\theta}=5$。

| 母體參數（紅球數 $\theta$） | 出現三紅球的機率 $L(\theta)$ |
|---|---|
| 3（3 紅 2 白） | $C_3^3/C_3^5=1/10$ |
| 4（4 紅 1 白） | $C_3^4/C_3^5=4/10$ |
| 5（5 紅） | $C_3^5/C_3^5=1$ |



● 特定隨機樣本出現的機率：假設有一母體 $X$，從母體中抽出一隨機樣本 $X_1,\cdots,X_n$，並令 $x_1,\cdots,x_n$ 為該隨機樣本的一個實現值。

- 間斷母體：若母體是間斷隨機變數，其機率函數以 $f(X;\theta)$ 表示，其中 $\theta$ 為未知母體參數[需要加以估計]。當隨機樣本的觀察值為 $x_1,\cdots,x_n$ 時，觀察到這些個別樣本的機率分別為 $f(x_1;\theta)$、...、$f(x_n;\theta)$，因此樣本觀察值 $x_1,\cdots,x_n$ 出現的機率為[聯合機率，運用隨機樣本相互獨立的特性]

$$L(x_1,\cdots x_n;\theta)=f(x_1;\theta)\cdots f(x_n;\theta)$$

由於 $x_1,\cdots,x_n$ 為已知，因此可將 $L(x_1,\cdots x_n;\theta)$ 視為 $\theta$ 的函數

$$L(\theta)=L(x_1,\cdots x_n;\theta)=f(x_1;\theta)\cdots f(x_n;\theta)$$

- 連續母體：若母體是連續隨機變數，其機率密度函數以 $f(X;\theta)$ 表示，其中 $\theta$ 為未知母體參數[需要加以估計]。當隨機樣本的觀察值為 $x_1,\cdots,x_n$ 時，觀察到這些個別樣本的機

率密度分別為 $f(x_1;\theta)$ 、...、 $f(x_n;\theta)$ ，因此樣本觀察值 $x_1,\cdots,x_n$ 出現的機率密度為[聯合機率密度，運用隨機樣本相互獨立的特性]

$$L(x_1,\cdots x_n;\theta)=f(x_1;\theta)\cdots f(x_n;\theta)$$

由於 $x_1,\cdots,x_n$ 為已知，因此可將 $L(x_1,\cdots x_n;\theta)$ 視為 $\theta$ 的函數。

$$L(\theta)=L(x_1,\cdots x_n;\theta)=f(x_1;\theta)\cdots f(x_n;\theta)$$

- 在間斷母體中， $L(x_1,\cdots x_n;\theta)$ 就是樣本觀察值 $x_1,\cdots,x_n$ 發生的機率，但連續母體中 $L(x_1,\cdots x_n;\theta)$ 是樣本觀察值 $x_1,\cdots,x_n$ 發生的機率密度，我們將其稱為**概似值**(likelihood)。若將 $L(x_1,\cdots x_n;\theta)$ 視為 $\theta$ 的函數時，則稱 $L(\theta)$ 為**概似函數**。
- 不管母體為間斷或連續，$L(x_1,\cdots x_n;\theta)$ 均為 $\theta$ 之函數；若 $\theta$ 的數值改變，樣本觀察值 $x_1,\cdots,x_n$ 的 $L(x_1,\cdots x_n;\theta)$ 亦隨之改變；而最大概似法的估計原理就是『找出使得 $L(x_1,\cdots x_n;\theta)$ 極大化的參數值 $\hat{\theta}$ 』。



- **概似函數(likelihood function)**：假設 $X_1,\cdots,X_n$ 為從分配 $f(X;\theta)$ 的母體中隨機抽出之一組隨機樣本，其中 $\theta$ 代表母體參數(未知，須加以估計)；若 $x_1,x_2,\cdots,x_n$ 為樣本觀察值，則概似函數為：

$$\begin{aligned}L(x_1,x_2,\cdots,x_n;\theta)&=f(x_1;\theta)f(x_2;\theta)....f(x_n;\theta)\\&=\prod_{i=1}^n f(x_i;\theta)\\&=L(\theta)\end{aligned}$$

要極大化 $L(\theta)$ 較複雜，因此通常將概似函數取自然對數。

**對數概似函數(log likelihood function)**

$$\begin{aligned}\ln L(\theta)&=\ln f(x_1;\theta)+\ln f(x_2;\theta)+\cdots+\ln f(x_n;\theta)\\&=\sum_{i=1}^n\ln f(x_i;\theta)\end{aligned}$$

[自然對數為單調遞增函數，若求出的 $\hat{\theta}$ 可以使得 $\ln L(\theta)$ 達到極大值，則 $\hat{\theta}$ 亦使得 $L(\theta)$ 達到極大值]



- **最大概似法**：找出估計式 $\hat{\theta}$ 使得概似函數達到極大值，亦即找出滿足底下條件的 $\hat{\theta}$ ：

$$\frac{\partial L(\theta)}{\partial\theta}=0 \quad 且 \quad \frac{\partial^2 L(\theta)}{\partial\theta^2}<0$$

或者是

$$\frac{\partial\ln L(\theta)}{\partial\theta}=0 \quad 且 \quad \frac{\partial^2\ln L(\theta)}{\partial\theta^2}<0$$

以最大概似法所求出的估計式稱為**最大概似估計式**(maximum likelihood estimator; MLE)

- 最大概似估計式的性質：
  - MLE 必為一致性估計式。
  - MLE 具漸近不偏性，但不一定具有不偏性。
  - MLE 具漸近有效性[樣本數較大時具相對有效性]。
  - MLE 需知道母體分配函數 $f(X;\theta)$ 。



- 例子：假設母體 $X$ 具有伯努利分配 $f(x;p)=p^x(1-p)^{1-x},x=0,1$ ；若從中抽出一組隨機樣本 $X_1,\cdots,X_n$ ，則 $p$ 之 MLE 為何？
  - 概似函數為

$$\begin{aligned}L(p)&=f(x_1;p)\cdots f(x_n;p)\\&=[p^{x_1}(1-p)^{1-x_1}][p^{x_2}(1-p)^{1-x_2}]\cdots[p^{x_n}(1-p)^{1-x_n}]\\&=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\end{aligned}$$

  - 對數概似函數為 $\ln L(p)=(\sum_{i=1}^n x_i)\ln p+(n-\sum_{i=1}^n x_i)\ln(1-p)$
  - 一階條件： $\ln L(p)$ 對 $p$ 的一階微分等於 0

$$\frac{\partial\ln L(p)}{\partial p}=(\sum_{i=1}^n x_i)\frac{1}{p}+(n-\sum_{i=1}^n x_i)\frac{-1}{(1-p)}=0$$

$$\Rightarrow\hat{p}=\tfrac{1}{n}\sum_{i=1}^n x_i$$

  - 因此 $p$ 之估計式為 $\Rightarrow\hat{p}=\frac{1}{n}\sum_{i=1}^n X_i$ [樣本比例為 MLE]

- 例子：假設有一母體 $X \sim N(\mu,\sigma^2)$，則 $\mu$ 與 $\sigma^2$ 的 MLE 為何？
  - 概似函數為

$$L(\mu,\sigma^2) = f(x_1;\mu,\sigma^2)\cdots f(x_n;\mu,\sigma^2)$$
$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$
$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}$$

  - 對數概似函數為

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2$$

  - 一階條件：

$\ln L(\mu,\sigma^2)$ 對 $\mu$ 的一階偏微分等於 0

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu) = 0$$



$$\Rightarrow \sum_{i=1}^n (x_i-\mu) = 0$$
$$\Rightarrow \sum_{i=1}^n x_i - n\mu = 0$$
$$\Rightarrow n\mu = \sum_{i=1}^n x_i$$
$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$$

[樣本平均數為母體平均數的 MLE]

$\ln L(\mu,\sigma^2)$ 對 $\sigma^2$ 的一階偏微分等於 0

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (x_i-\mu)^2 = 0$$
$$\Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i-\mu)^2 = 0$$
$$\Rightarrow n\sigma^2 = \sum_{i=1}^n (x_i-\mu)^2$$
$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{X})^2$$

[在估計變異數時除以樣本個數 $n$ 的公式為 MLE]

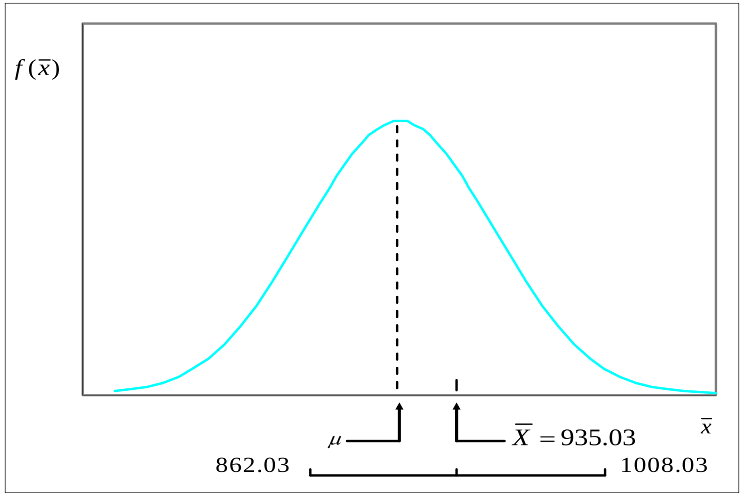


## 區間估計(Interval estimate)

- 我們已經瞭解好的估計式應具備什麼性質，也介紹過如何找出優良估計式的一般性方法。
- 但是，即使你已經找到一個優良的估計式，並將該估計式運用到一組隨機樣本計算出一個估計值，該估計值也無法告訴我們『估計的精準度如何』
  - 例子：王太太估計出『台北市區 25~30 坪的房屋平均價格為 935.03 萬元』，雖然我們知道 935.03 萬元這個估計值是運用樣本平均數(好的估計式)算出，但這個數字並無法回答底下問題：『這個估計值多可靠？』、『估計誤差多大？』
  - 點估計值無法提供推論母體參數時之可靠度與準確度相關訊息；若我們在估計時要將有關精準度的訊息揭露出來，較好的方法是估計出一個區間，並指出該區間包含母體參數的可靠度(機率)有多高。



- **區間估計**：對未知的母體參數估計出一個上下限的區間，並指出該區間包含母體參數的可靠度(機率)。
  - 例子：繼續房屋平均價格的例子。若以 $\bar{X} = 935.03$ 為中心加減某個數字(例如 73)，得出 $[935.03-73, 935.03+73]$ 的區間([862.03,1008.03])，若我們可以用某種方法[估計式的抽樣分配]推論出真實房價介於此區間的機率(例如 95%)，則我們可以說『我們有 95%的信心，台北市區 25~30 坪的房屋平均價格會介於 862.03 ~ 1008.03 萬元』，這就是一個區間估計(值)。
  - 區間估計的作法：由於估計式一定具有抽樣分配，因此我們可以算出估計值位於某個區間的機率。區間估計即是由點估計值出發，以點估計值為中心加減某個數字，若根據估計式的抽樣分配計算出估計值位於此區間的機率為 $1-\alpha$ [信賴水準]，則稱此區間為『$(1-\alpha)$100%信賴區間』。




- 實際區間估計 vs.近似(漸近)區間估計：由於抽樣分配有實際分配與漸近分配之區分，因此信賴區間也會根據抽樣分配之推導立論不同而有所區分。



- 進一步說明：估計式是具有某種機率分配(抽樣分配)的隨機變數，而估計值只是該隨機變數的一個觀察值(實現值)，該觀察值不見得接近我們有興趣的真實參數。因此，只提供點估計無法告知任何跟估計準確度有關的資訊。
  - 例如：以樣本平均數 550 來做為為母體平均數的點估計值，該數字(550)並未告訴我們母體平均數有多靠近這個估計值
  - 然而，若我們說『我們有 99%的信心母體平均數會落入區間 [449,511]』，這句話傳達了更多有關母體平均數的資訊
  - 若有另一區間『我們有 90%的信心母體平均數會落入區間 [400,700]』，與這個區間相比較，第一個區間估計相對較準確[不只信心度較高、且區間較窄]。
- **信賴區間(confidence interval)**：信賴區間是在一個既定的信賴水準下所構成的一個區間。是由樣本統計量及抽樣誤差所構成的一個(包含上限、下限的)區間。
- **信賴水準(level of confidence)[信賴係數]**：信賴水準是指信賴區間包含母體參數的信心(或稱可靠度，信賴度)。



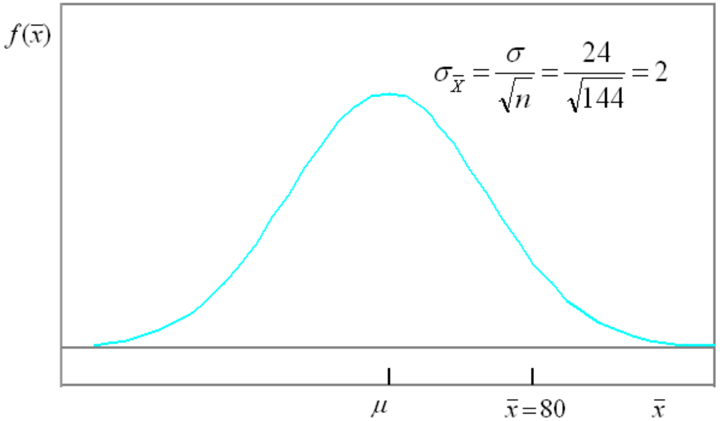
## 母體平均數的區間估計—非常態母體且變異數已知(11.4.1)

- 非常態母體樣本平均數之抽樣分配：假設隨機樣本 $X_1,\cdots,X_n$ 來自於平均數為 $\mu$ 、變異數為 $\sigma^2$ 的(非常態)母體；令 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 為樣本平均數。我們已知 $\mu_{\bar{X}} = E(\bar{X}) = \mu$ 、 $\sigma_{\bar{X}}^2 = V(\bar{X}) = \frac{\sigma^2}{n}$ ，因此標準化的樣本平均數為 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ 。
  根據中央極限定理，當樣本個數 $n \to \infty$ 時，標準化的樣本平均數具有標準常態分配，亦即
  $$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{n\to\infty} N(0,1)$$
  因此，當樣本數 $n$ 夠大時， $\bar{X}$ 具有底下的近似分配
  $$\bar{X} \approx N(\mu, \tfrac{\sigma^2}{n}) \qquad 或 \qquad \tfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$$



- 現在假設(非常態)母體平均數 $\mu$ 未知、需要估計，但母體變異數 $\sigma^2$ 已知的情況下，以下例說明如何進行區間估計[近似區間]。
- 例子：某選區一位立法委員想了解選民對其平時服務的滿意程度，因此從其選區之選民中隨機抽出 144 人[樣本個數 $n=144$ ]進行問卷調查，抽樣調查的結果顯示滿意程度的平均分數為 80 分( $\bar{X}=80$ )。而根據過去經驗，滿意程度的變異數相當穩定，約為 576 [ $\sigma^2=576$ ]，則如何建構區間估計呢？
  - **步驟 1**：選擇較佳的點估計式並計算點估計值。[優良的點估計式可使區間估計在相同的信心水準下區間範圍較窄]
    - 此例當然是利用樣本平均數來估計母體平均數。而我們已知樣本平均數的點估計值為 $\bar{X}=80$ 。
  - **步驟 2**：取得樣本統計量的抽樣分配。[抽樣分配描述的是抽樣誤差的分配情況，而取得抽樣分配的主要目的在於：決定樣本統計量之抽樣誤差介於某個區間範圍內的機率]

- 只要樣本個數 $n$ 夠大，非常態母體樣本平均數之分配近似常態，亦即 $\overline{X} \approx N(\mu, \frac{\sigma^2}{n})$；又因 $\sigma^2 = 576$，所以 $\sigma_{\overline{X}}^2 = \sigma^2 / n = 576/144 = 4$。該抽樣分配如下圖所示，而樣本平均數 $\overline{X} = 80$ 也許落在圖上所標示的位置。



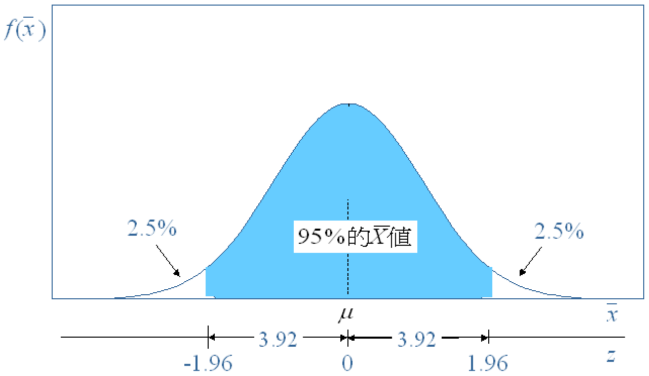



- 已知 $\overline{X} \approx N(\mu, \frac{\sigma^2}{n})$，若令 $\sigma_{\overline{X}}^2 = \sigma^2 / n$，則

$$\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \approx N(0,1)$$

查常態分配表我們可輕易得知 $\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$ 介於某區間範圍內的機率，例如我們可以知道

$$P\left(-1.96 \le \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \le 1.96\right) = 95\%$$






該式可改寫為

$$P\left(\left|\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}\right| \le 1.96\right) = 95\% \Rightarrow P(|\overline{X} - \mu| \le 1.96\sigma_{\overline{X}}) = 0.95$$

抽樣誤差 $= |\overline{X} - \mu|$ 小於等於 $1.96\sigma_{\overline{X}}$ 的機率有 0.95 [此即樣本統計量抽樣誤差介於某個區間範圍內的機率]

上式可進一步改寫為

$$P(\mu - 1.96\sigma_{\overline{X}} \le \overline{X} \le \mu + 1.96\sigma_{\overline{X}}) = 0.95$$

[$\overline{X}$ 介於母體平均數上下 $1.96\sigma_{\overline{X}} = 3.92$ 的機率為 0.95]

- 符號補充：令常態隨機變數大於 $Z_{\alpha/2}$ 的機率為 $\alpha/2$ [右尾機率]，再加上常態分配具對稱的性質，可得

$$P\left(-Z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \le Z_{\alpha/2}\right) = 1-\alpha$$

故抽樣誤差介於某個區間內的一般化表達式為

$$P(|\overline{X} - \mu| \le Z_{\alpha/2}\sigma_{\overline{X}}) = 1-\alpha$$



➢ **步驟 3**：導出母體參數的信賴區間。[利用抽樣誤差的機率分配，反推母體參數在某個機率水準下應介於哪個範圍]

- 雖然已知 $P(\mu - 1.96\sigma_{\overline{X}} \le \overline{X} \le \mu + 1.96\sigma_{\overline{X}}) = 0.95$，但我們的目的是估計母體平均數 $\mu$ 介於哪個範圍。將上式進行些許代數運算即可得到我們想要的結果

$$P(\mu - 1.96\sigma_{\overline{X}} \le \overline{X} \le \mu + 1.96\sigma_{\overline{X}}) = 0.95$$

$$\Rightarrow P(-1.96\sigma_{\overline{X}} \le \overline{X} - \mu \le 1.96\sigma_{\overline{X}}) = 0.95$$

$$\Rightarrow P(-1.96\sigma_{\overline{X}} \le \mu - \overline{X} \le 1.96\sigma_{\overline{X}}) = 0.95$$

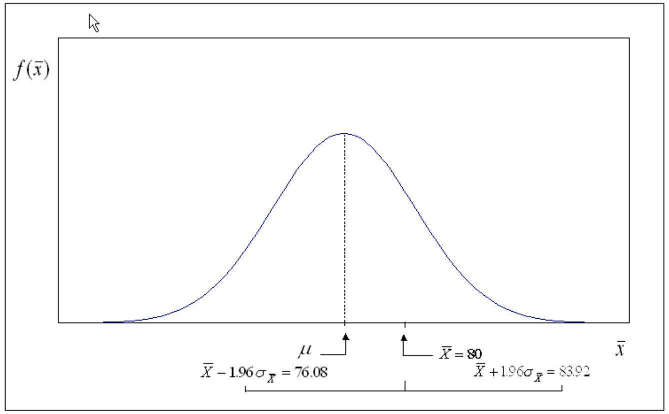
$$\Rightarrow P(\overline{X} - 1.96\sigma_{\overline{X}} \le \mu \le \overline{X} + 1.96\sigma_{\overline{X}}) = 0.95$$

這就是『我們有 95%的信心，母體平均數會介於 $[\overline{X} - 1.96\sigma_{\overline{X}}, \overline{X} + 1.96\sigma_{\overline{X}}]$ 這個區間內』。$\overline{X} - 1.96\sigma_{\overline{X}}$ 為區間下限，$\overline{X} + 1.96\sigma_{\overline{X}}$ 為區間上限。

- 信賴區間[**信賴區間估計式**]的其他表達方式：

$$(\overline{X} - 1.96\sigma_{\overline{X}}) \sim (\overline{X} + 1.96\sigma_{\overline{X}})\text{、}\overline{X} \pm 1.96\sigma_{\overline{X}}$$

➢ **步驟 4**：求出母體參數的信賴區間值並做統計推論。[將樣本所得的資訊代入區間估計式中，計算區間估計值]




■ 母體平均數的 95%信賴區間估計式為 $\bar{X}\pm1.96\sigma_{\bar{X}}$，將 $\bar{X}=80$、$\sigma_{\bar{X}}=2$ 代入可得 95%信賴區間估計值

$$\bar{X}\pm1.96\sigma_{\bar{X}}=80\pm1.96\cdot2$$



亦即母體平均數在 95%的信心水準下介於底下範圍

$$76.08\le\mu\le83.92$$

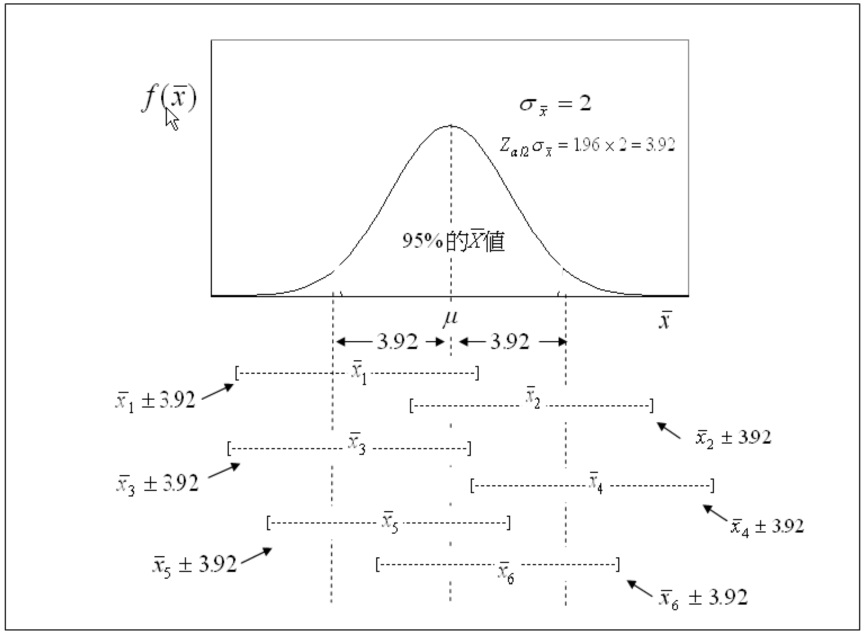
因此我們可以做出底下結論，並運用該結論進行決策『在 95%的信賴水準下，服務滿意度分數介於 76.08 分~83.92 分之間』[註：這是一個近似區間]

● 95%信賴區間的意義：

➢ 『$\mu$ 的 95%信賴區間 $\bar{X}\pm1.96\sigma_{\bar{X}}$』的意義並不是『$\mu$ 落在區間 $\bar{X}\pm1.96\sigma_{\bar{X}}$ 的機率是 95%』。因為母體平均數 $\mu$ 並非隨機變數(而是固定常數)，因此 $\mu$ 落在區間 $\bar{X}\pm1.96\sigma_{\bar{X}}$ 的機率不是 0 就是 1，絕非 95%。

➢ 若站在抽樣前的角度來看，$\bar{X}$ 為隨機變數，因此 $\bar{X}\pm1.96\sigma_{\bar{X}}$ 事實上是**隨機區間**(random interval)。而 95%信賴區間的真實意義是『每次抽樣後都以 $\bar{X}\pm1.96\sigma_{\bar{X}}$ 建構 $\mu$ 的區間估計值，假設我們可以抽樣 100 次，建構出 100 個 $\mu$ 的區間估計值，則這 100 個區間約有 95 個會包含母體平均數 $\mu$』








● **小結**：非常態母體，母體變異數 $\sigma^2$ 已知，母體平均數 $(1-\alpha)100\%$ 信賴區間為：[其中 $\sigma_{\bar{X}}=\sigma/\sqrt{n}$]

$$\bar{X}\pm Z_{\alpha/2}\sigma_{\bar{X}}$$

➢ $\bar{X}-Z_{\alpha/2}\sigma_{\bar{X}}$ 稱為信賴區間下限，$\bar{X}+Z_{\alpha/2}\sigma_{\bar{X}}$ 為信賴區間上限

➢ 若我們所設定的信賴水準 $(1-\alpha)$ 改變，則 $Z_{\alpha/2}$ 隨之改變，因此信賴區間亦有所不同。下表整理出各種信賴水準 $(1-\alpha)$ 下的母體平均數區間估計式：[統計學慣用 0.90、0,95、0.99 等信賴水準，亦即 $\alpha=0.10,0.05,0.01$]

| 信賴水準 $1-\alpha$ | $\alpha$ | $\alpha/2$ | $Z_{\alpha/2}$ | 信賴區間 $\bar{X}\pm Z_{\alpha/2}\sigma_{\bar{X}}$ |
|---|---|---|---|---|
| 0.90 | 0.10 | 0.05 | 1.645 | $\bar{X}\pm1.645\sigma_{\bar{X}}$ |
| 0.95 | 0.05 | 0.025 | 1.96 | $\bar{X}\pm1.96\sigma_{\bar{X}}$ |
| 0.99 | 0.01 | 0.005 | 2.575 | $\bar{X}\pm2.575\sigma_{\bar{X}}$ |

- 例子：青少年上網時間[完整的答題邏輯]

問題：尋找青少年每週上網時間的 95%信賴區間。

已知：隨機抽出台中市 64 位青少年調查其上網時間，結果發現 64 位青少年每週上網時間的平均值為 $\bar{X}=15.6$ 小時，並假設已知每週上網時間的標準差為 $\sigma=5$

**解答：**令 $\mu$ 代表青少年上網時間的(母體)平均數，$\sigma=5$ 代表上網時間的標準差。由於每週上網時間的分配未知，因此無法得知樣本平均數 $\bar{X}$ 的實際分配；但根據中央極限定理，當樣本個數 $n$ 夠大，標準化的樣本平均數近似標準常態分配 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$

由常態分配的機率表可知 $P(-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96)=0.95$，因此

$$P(-1.96\cdot\sigma/\sqrt{n} \le \bar{X}-\mu \le 1.96\cdot\sigma/\sqrt{n})=0.95$$

$$\Rightarrow P(-1.96\cdot\sigma/\sqrt{n} \le \mu-\bar{X} \le 1.96\cdot\sigma/\sqrt{n})=0.95$$

$$\Rightarrow P(\bar{X}-1.96\cdot\sigma/\sqrt{n} \le \mu \le \bar{X}+1.96\cdot\sigma/\sqrt{n})=0.95$$

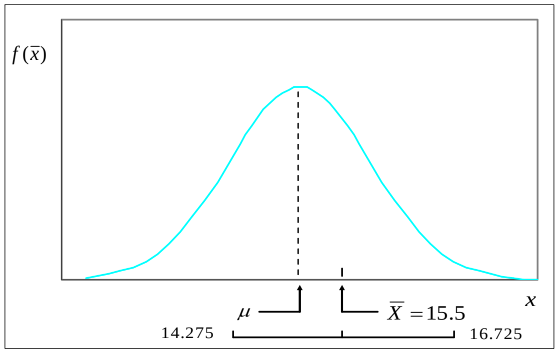


故母體平均數 $\mu$ 的 95%信賴區間估計式為 $\bar{X}\pm1.96\cdot\sigma/\sqrt{n}$。將樣本平均數 $\bar{X}=15.6$、樣本個數 $n=64$、母體標準差 $\sigma=5$ 代入可得青少年每週上網時間的 95%信賴區間估計值為

$$\bar{X}\pm1.96\cdot\sigma/\sqrt{n}=15.6\pm1.96\cdot5/\sqrt{64}=15.6\pm1.225$$

$$=14.275\sim16.725$$

『青少年每週上網時間的 95%信賴區間為 $14.275\sim16.725$ 小時』






- **信賴區間長度的決定因素：**
  - 若我們用兩種方法估計出的 95%信賴區間分別為[2,3]與[1,4]，直覺上會認為[2,3]的估計較準確(因區間長度較小)
  - 區間估計 $\bar{X}\pm Z_{\alpha/2}\sigma_{\bar{X}}$ 的長度 $2Z_{\alpha/2}\sigma_{\bar{X}}$ 取決於 $Z_{\alpha/2}$ 與 $\sigma_{\bar{X}}$ 的大小，在給定的信賴水準 $1-\alpha$ 下 $Z_{\alpha/2}$ 為一固定常數，故信賴區間之長度主要取決於估計式的標準差 $\sigma_{\bar{X}}$，而 $\sigma_{\bar{X}}$ 越小即代表估計約準確。
  - 若有一區間估計方法，其所估計的信賴區間長度(相對於其他區間估計方法而言)較小，就是較準確的區間估計方法。
  - 底下詳細說明影響信賴區間長度的因素：

(1) 所選擇的點估計式的抽樣分配：較優良的(不偏)估計式具有較小的抽樣分配標準差 $\sigma_{\bar{X}}$，因此運用較優良的點估計式所找出的信賴區間長度較小[這就是為什麼區間估計的第一個步驟是要選擇一個優良估計式的主要原因]。



(2) 樣本數：因為抽樣分配的標準差與樣本個數 $n$ 有關，亦即 $\sigma_{\bar{X}}=\sigma/\sqrt{n}$，因此當樣本個數 $n$ 越大時，抽樣分配的標準差 $\sigma_{\bar{X}}$，信賴區間長度亦較小。

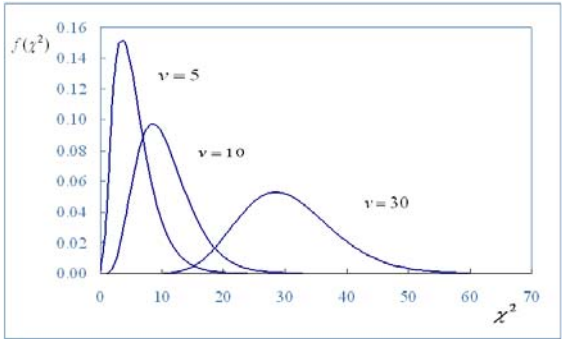
(3) 機率區間上下限的取法：給定信賴水準 $1-\alpha$ 下，進行區間估計 $\bar{X}\pm Z_{\alpha/2}\sigma_{\bar{X}}$ 時[將點估計值加減 $Z_{\alpha/2}$ 倍的估計式標準差]，我們使用唯一一個 $Z_{\alpha/2}$；這等於是在常態分配的兩個尾端各分配 $\alpha/2$ 的機率。會如此做的主要原因是『創造區間長度較小的信賴區間』[若估計式的抽樣分配是對稱的，偏離此作法將會造成估計區間長度增大]

(4) 信賴水準：信賴水準 $1-\alpha$ 改變時，$Z_{\alpha/2}$ 即會改變；信賴水準 $1-\alpha$ 越大，$Z_{\alpha/2}$ 越大，估計區間長度亦越大。

- 以上說明雖僅針對母體平均數的區間估計式，但其實適用到其他母體參數的區間估計問題。

## 常態母體變異數的區間估計(11.8)

- 卡方(chi-square; $\chi^2$)分配：令 $Z_1, Z_2, ..., Z_v$ 為 $v$ 個獨立的標準常態隨機變數，且令 $W = Z_1^2 + \cdots Z_v^2 = \sum_{i=1}^{v} Z_i^2$，則稱 $W$ 具有<u>自由度為 $v$ 的卡方分配</u>。通常以 $W \sim \chi^2(v)$ 或 $W \sim \chi_v^2$ 來表示。
  - 卡方分配為一定義在大於等於 0 (正數)範圍的右偏分配，不同的自由度 $v$ 決定不同的卡方分配。

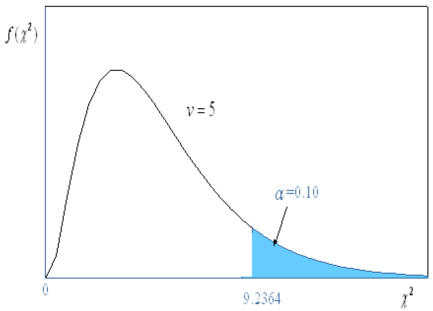


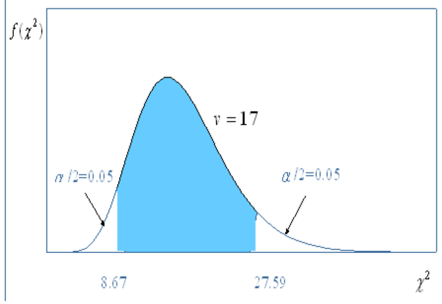



- 在課表後面 p.785~786 的表六中，列出了各種自由度下滿足 $P(\chi^2 > \chi_\alpha^2) = \alpha$ 的 $\chi_\alpha^2$ 值。

  **例如：**表中 $v = 5$ 與 $\chi_{0.10}^2$ 交叉的數字 9.23635 表示，若 $W$ 是自由度為 5 的卡方分配，則 $P(W > 9.23635) = 0.10$ [左圖]

  **例如：**表中 $v = 17$ 與 $\chi_{0.95}^2$ 交叉的數字 8.671754 表示，若 $W$ 是自由度為 17 的卡方分配，則 $P(W > 8.671754) = 0.95$，或是說 $P(W < 8.671754) = 0.05$ [右圖]






- 卡方分配只有一個參數，即其<u>自由度</u>(degree of freedom) $v$。
- 當自由度增加時，卡方分配會趨於對稱；當自由度趨近無窮大時($v \to \infty$)，<u>標準化的卡方分配會趨近於標準常態分配</u>
- 假設 $X \sim N(\mu, \sigma^2)$，令 $Z^2 = (\frac{X-\mu}{\sigma})^2$，則 $Z^2$ 為自由度 1 的卡方分配。
- 若 $W \sim \chi^2(v)$，則 $W$ 可表達為 $v$ 個獨立標準常態隨機變數的平方和。
- 卡方分配的可加性：若 $W_1 \sim \chi^2(m)$、$W_2 \sim \chi^2(n)$，且 $W_1$ 與 $W_2$ 互相獨立，則 $W_1 + W_2 \sim \chi^2(m+n)$
- 卡方分配的平均數與變異數為：$E(\chi_v^2) = v$、$\mathrm{var}(\chi_v^2) = 2v$

  **證明：**令 $W \sim \chi^2(v)$，則 $W = Z_1^2 + \cdots Z_v^2 = \sum_{i=1}^{v} Z_i^2$，其中 $Z_i$ 為獨立標準常態隨機變數。我們首先證明幾個有關 $Z_i$ 的性質

  (1) $E(Z_i^2) = E[(Z_i - 0)^2] = \mathrm{var}(Z_i) = 1$



(2) 若 $i \neq j$，則 $Z_i$ 與 $Z_j$ 獨立，因此 $Z_i^2$ 與 $Z_j^2$ 亦獨立，故其共變異數等於零：$\mathrm{cov}(Z_i^2, Z_j^2) = 0$。因此

$$E(Z_i^2 Z_j^2) = \mathrm{cov}(Z_i^2, Z_j^2) + E(Z_i^2)E(Z_j^2) = 0 + 1 \cdot 1 = 1$$

(3) $E(Z_i^4) = 3$

$$E(Z_i^4) = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} -x^3 d(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}})$$

$$= -x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 3x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 3$$

運用結果(2)與(3)可知

$$E[(\chi_v^2)^2] = E(W^2) = E[(\sum_{i=1}^{v} Z_i^2)^2]$$

$$= E[\sum_{i=1}^{v} Z_i^4 + 2\sum_{i=1}^{v}\sum_{j=i+1}^{v} Z_i^2 Z_j^2]$$

$$= \sum_{i=1}^{v} E(Z_i^4) + 2\sum_{i=1}^{v}\sum_{j=i+1}^{v} E(Z_i^2 Z_j^2)$$

$$= \sum_{i=1}^{v} 3 + 2\sum_{i=1}^{v}\sum_{j=i+1}^{v} 1 = 3v + v(v-1)$$

運用上述結果可知

$$E(\chi_v^2)=E(W)=E(\textstyle\sum_{i=1}^{v}Z_i^2)=\sum_{i=1}^{v}E(Z_i^2)=\sum_{i=1}^{v}1=v$$

$$\mathrm{var}(\chi_v^2)=E[(\chi_v^2)^2]-[E(\chi_v^2)]^2=3v+v(v-1)-v^2=2v$$

- 常態母體變異數的估計問題：假設母體 $X\sim N(\mu,\sigma^2)$，其中 $\mu$ 與 $\sigma^2$ 均是未知參數，從母體中取得隨機樣本 $X_1,...,X_n$。在估計母體變異數時以樣本變異數為之，

$$S^2=\tfrac{1}{n-1}\textstyle\sum_{i=1}^{n}(X_i-\bar{X})^2$$

- 常態母體樣本變異數的抽樣分配：樣本變異數本身並沒有分配，但若將樣本變異數乘以 $\frac{n-1}{\sigma^2}$，亦即

$$\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$$

則 $\frac{(n-1)S^2}{\sigma^2}$ 具有自由度 $(n-1)$ 的卡方分配，亦即 $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$



證明：首先證明『$\bar{X}$ 與 $S^2$ 獨立』：若可證明 $\bar{X}$ 與 $X_i-\bar{X}$ 獨立，則 $\bar{X}$ 與 $(X_i-\bar{X})^2$ 亦獨立，因此 $\bar{X}$ 與 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 亦獨立。所以我們只需證明『$\bar{X}$ 與 $X_i-\bar{X}$ 獨立』即可，但由於

$$\bar{X}=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}X_i \qquad \text{[常態隨機變數的線性組合]}$$

$$X_i-\bar{X}=X_i-\tfrac{1}{n}\textstyle\sum_{j=1}^{n}X_j=X_i-\tfrac{1}{n}X_i-\tfrac{1}{n}\sum_{j\neq i}X_j$$

$$=(1-\tfrac{1}{n})X_i-\tfrac{1}{n}\textstyle\sum_{j\neq i}X_j \qquad \text{[常態隨機變數的線性組合]}$$

因此 $\bar{X}$ 與 $X_i-\bar{X}$ 均為常態隨機變數，故僅需證明 $\bar{X}$ 與 $X_i-\bar{X}$ 之共變異數為 0 即可。由於

$$\mathrm{cov}(\bar{X},X_i-\bar{X})=\mathrm{cov}(\bar{X},X_i)-\mathrm{cov}(\bar{X},\bar{X})=\mathrm{cov}(\bar{X},X_i)-\mathrm{var}(\bar{X})$$

而 $\mathrm{cov}(\bar{X},X_i)=\mathrm{cov}(\frac{1}{n}\sum_{j=1}^{n}X_j,X_i)=\frac{1}{n}\sum_{j=1}^{n}\mathrm{cov}(X_j,X_i)$

$$=\tfrac{1}{n}\mathrm{cov}(X_i,X_i)=\tfrac{1}{n}\mathrm{var}(X_i)=\mathrm{var}(\bar{X})$$

因此 $\mathrm{cov}(\bar{X},X_i-\bar{X})=0$



接下來我們可以得到

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2\right]$$

$$=\frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{n}(\bar{X}-\mu)^2+2\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu)\right]$$

但因 $\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu)=(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X})=0$

而 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\Rightarrow\sum_{i=1}^{n}(X_i-\bar{X})^2=(n-1)S^2$

所以 $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\dfrac{(n-1)S^2}{\sigma^2}+\dfrac{\sum_{i=1}^{n}(\bar{X}-\mu)^2}{\sigma^2}$

或者是{式中用到 $\sum_{i=1}^{n}(\bar{X}-\mu)^2=n(\bar{X}-\mu)^2=[\sqrt{n}(\bar{X}-\mu)]^2$}

$$\boxed{\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2-\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2}$$



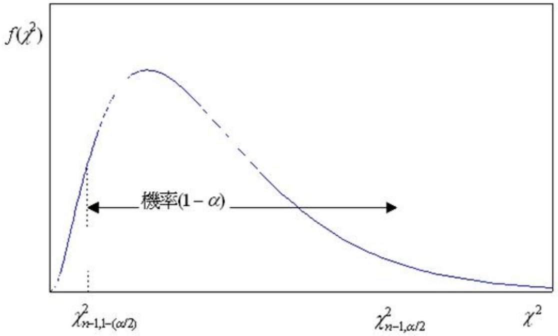
由於 $\dfrac{X_i-\mu}{\sigma}\sim N(0,1)$、$\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$

故 $\displaystyle\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$、$\left(\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2\sim\chi^2(1)$

因為兩獨立的卡方隨機變數具可加性(因 $\bar{X}$ 與 $S^2$ 獨立，故 $\dfrac{(n-1)S^2}{\sigma^2}$ 與 $\left(\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2$ 獨立)

$$\frac{(n-1)S^2}{\sigma^2}+\underbrace{\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2}_{\chi^2(1)}=\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\chi^2(n)}$$

因此 $\boxed{\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)}$




- **母體變異數的信賴區間**：假設給定信賴水準$1-\alpha$，將$\alpha$的機率平分給兩個尾端(每邊各$\alpha/2$)，由於$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，我們可根據$\chi^2(n-1)$分配找出$\chi^2_{n-1,1-\alpha/2}$與$\chi^2_{n-1,\alpha/2}$，因此

$$P\left(\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,\alpha/2}\right) = 1-\alpha$$



左邊不等式：$\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \Rightarrow \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}$

右邊不等式：$\frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,\alpha/2} \Rightarrow \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2$

合併：$P\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$

因此常態母體變異數$\sigma^2$的$(1-\alpha)100\%$信賴區間為

$$\boxed{\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}}$$

- 而常態母體標準差$\sigma$的$(1-\alpha)100\%$信賴區間為

$$\boxed{\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}}}$$



- 例子：某進口商抽驗 20 罐貓食罐頭，發現每罐容量的標準差為 25g，請問每罐容量的母體變異數及標準差的 95%信賴區間？
  - 題目雖沒宣稱母體為常態分配，但若不利用上述方法已經沒其他方法可用了。
  - 由於$n=20$，必須根據$\chi^2(19)$分配找出左右尾<u>臨界值</u>$\chi^2_{n-1,1-\alpha/2}$與$\chi^2_{n-1,\alpha/2}$，亦即$\chi^2_{19,0.925}$與$\chi^2_{19,0.025}$。查表可知$\chi^2_{19,0.925}=8.91$、$\chi^2_{19,0.025}=32.85$，因此母體變異數的 95%信賴區間為

$$\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \Rightarrow \frac{(20-1)25^2}{32.85} \le \sigma^2 \le \frac{(20-1)25^2}{8.91}$$

$$\Rightarrow 361.49 \le \sigma^2 \le 1332.77$$

  - 母體標準差的 95%信賴區間為

$$\sqrt{361.49} \le \sigma \le \sqrt{1332.77} \Rightarrow 19.01 \le \sigma \le 36.51$$



- 欠大家的一個證明：$V(S^2)=\frac{2\sigma^4}{n-1}$

由於$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，因此$V\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$

故$\frac{(n-1)^2}{\sigma^4}V(S^2) = 2(n-1) \Rightarrow V(S^2) = \frac{2\sigma^4}{n-1}$

## 母體平均數的區間估計——常態母體且變異數已知(11.5.1)

- 假設隨機樣本$X_1,\cdots,X_n$來自於平均數為$\mu$、變異數為$\sigma^2$的常態母體，且母體變異數$\sigma^2$已知，則樣本平均數具有底下抽樣分配(不管樣本個數$n$是多少)

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

$$\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- 若常態隨機變數大於 $Z_{\alpha/2}$ 的機率為 $\alpha/2$，則

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

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

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

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

因此，常態母體、母體變異數已知情況下，母體平均數的 $(1-\alpha)100\%$ 信賴區間為 $\boxed{\bar{X}\pm Z_{\alpha/2}\cdot\sigma/\sqrt{n}}$

- 例子：課本例 11.8 請自行演算(與非常態母體、變異數已知類似)

## 母體平均數的區間估計——常態母體但變異數未知(11.5.2)

- 延續上述的區間估計問題：若母體變異數 $\sigma^2$ 未知，在進行信賴區間 $\bar{X}\pm Z_{\alpha/2}\cdot\sigma/\sqrt{n}$ 的計算時，就須對母體變異數進行估計。我們很自然會以樣本變異數 $S^2$ 估計 $\sigma^2$ [即以 $S$ 估計 $\sigma$]，但區間估計可不是直接取代成 $\bar{X}\pm Z_{\alpha/2}\cdot S/\sqrt{n}$ 問題就解決了。

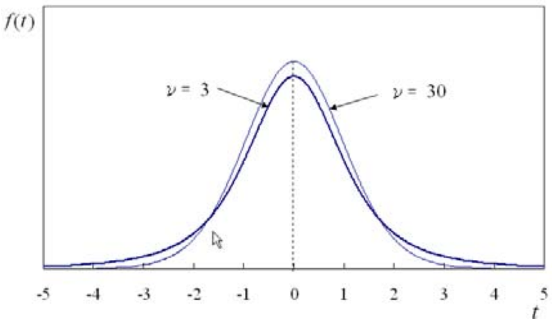


- $t$ 分配($t$ distribution)：令兩隨機變數 $Z\sim N(0,1)$、$W\sim\chi^2(v)$，且 $Z$ 與 $W$ 獨立，定義一新的隨機變數 $U$

$$U=\frac{Z}{\sqrt{\frac{W}{v}}}$$

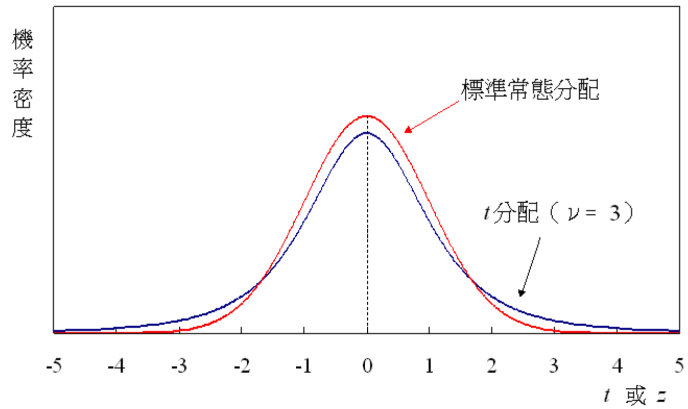
則 $U$ 為具有自由度 $v$ 的 $t$ 分配，通常以 $U\sim t(v)$ 或 $U\sim t_v$ 表示。

  - $t$ 分配為一個以平均數 0 為中心的對稱分配，不同的自由度 $v$ 有不同的 $t$ 分配。自由度 $v$ 為 $t$ 分配的唯一參數。






- $t$ 分配的平均數與變異數為：$E(t_v)=0$，$V(t_v)=\frac{v}{v-2}$
- 自由度趨近於無窮大時($v\to\infty$)，$t$ 分配趨近於標準常態分配，即 $t_v\approx N(0,1)$。若 $v\ge 30$，則 $t$ 分配非常接近常態分配。
- $t$ 分配曲線比標準常態曲線為平坦，亦即 $t$ 分配曲線的高度較低，分散程度較大；$t$ 分配的實現值範圍介於 $\pm\infty$ 之間。



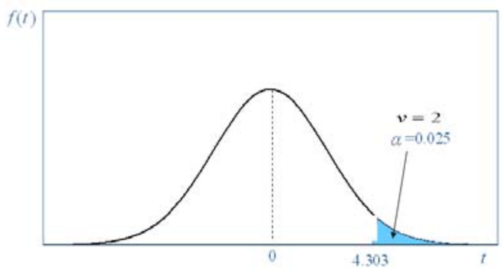


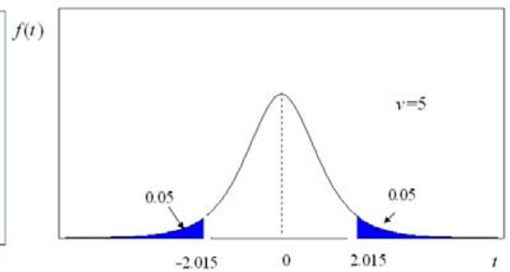

- $t$ 分配的臨界值(critical value)：課本後面 p.783~784 的表五列出不同自由度的 $t$ 分配滿足 $P(t>t_\alpha)=\alpha$ 的 $t_\alpha$ 值。

例如：對應到自由度 $v=2$ 與 $t_{.025}$ 的數字 4.303 代表，若 $U\sim t_2$，則 $P(U>4.303)=0.025$。[左圖]

例如：對應到自由度 $v=5$ 與 $t_{.05}$ 的數字 2.015 代表，若 $U\sim t_5$，則 $P(U>2.015)=0.05$。由於 $t$ 分配具對稱性，因此 $P(U<-2.015)=0.05$，而且[右圖]

$$P(-2.015<U<2.015)=0.90$$



- 常態母體、變異數$\sigma^2$未知下，樣本平均數的抽樣分配：假設你使用 $n$ 個樣本觀察值估計了樣本平均數$\bar{X}$與樣本變異數$S^2$，則

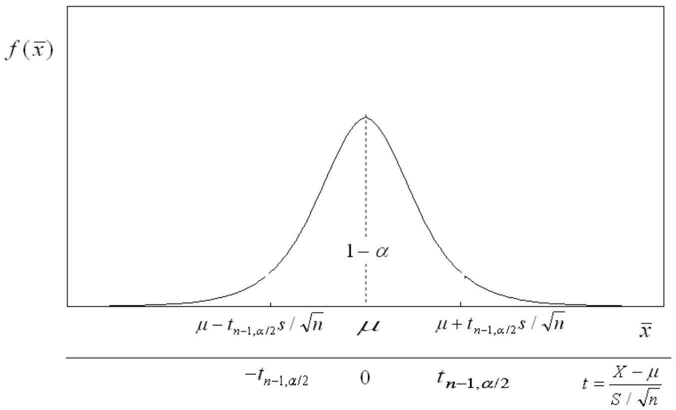
$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

**推導：**將$\frac{\bar{X}-\mu}{S/\sqrt{n}}$改寫如下

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} = \frac{\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}}$$

由於$\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$、$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，且$\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$與$\frac{(n-1)S^2}{\sigma^2}$獨立(因$\bar{X}$與$S^2$獨立)，故$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$。






- 常態母體、變異數未知時，母體平均數的信賴區間：假設給定信賴水準$1-\alpha$，將$\alpha$的機率平分給兩個尾端(每邊各$\alpha/2$)，根據$t(n-1)$分配找出臨界值$t_{n-1,\alpha/2}$，則

$$P\left(-t_{n-1,\alpha/2} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{n-1,\alpha/2}\right) = 1-\alpha$$



$$\Rightarrow P(-t_{n-1,\alpha/2}S/\sqrt{n} \le \bar{X}-\mu \le t_{n-1,\alpha/2}S/\sqrt{n}) = 1-\alpha$$

$$\Rightarrow P(-\bar{X}-t_{n-1,\alpha/2}S/\sqrt{n} \le -\mu \le -\bar{X}+t_{n-1,\alpha/2}S/\sqrt{n}) = 1-\alpha$$

$$\Rightarrow P(\bar{X}-t_{n-1,\alpha/2}S/\sqrt{n} \le \mu \le \bar{X}+t_{n-1,\alpha/2}S/\sqrt{n}) = 1-\alpha$$

因此，常態母體、變異數未知時，母體平均數的$(1-\alpha)100\%$信賴區間為$\boxed{\bar{X} \pm t_{n-1,\alpha/2}S/\sqrt{n}}$

- 例子：某政府單位進行一項文書處理訓練計畫，想了解公務人員完成訓練計畫的平均時間，假設訓練時間服從常態分配。隨機抽取 25 人，發現其平均訓練時間為 6 小時，標準差為 2 小時，建構平均訓練時間的 95%信賴區間
  - 令 $X$ 代表訓練時間，由題意知$n=25$、$\bar{X}=6$、$S=2$。因此平均訓練時間的 95%信賴區間為

$$\bar{X} \pm t_{n-1,\alpha/2}S/\sqrt{n} = 6 \pm 2.064 \cdot 2/\sqrt{25} = 6 \pm 0.8256$$



## 母體平均數的區間估計——非常態母體但變異數未知

- 若母體並非常態，則$\frac{\bar{X}-\mu}{S/\sqrt{n}}$並不會具有$t(n-1)$的實際分配，僅具有漸近常態分配。

**推導：**將$\frac{\bar{X}-\mu}{S/\sqrt{n}}$改寫為$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\frac{\sigma}{S}$，由於$S$為$\sigma$的一致性估計式，$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$具漸近常態，因此在樣本個數$n \to \infty$時，$\frac{\bar{X}-\mu}{S/\sqrt{n}} \to \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

- 非常態母體、變異數未知時，母體平均數的區間估計：給定信賴水準$1-\alpha$，則

$$P\left(-Z_{\alpha/2} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le Z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P(\bar{X}-Z_{\alpha/2} \cdot S/\sqrt{n} \le \mu \le \bar{X}+Z_{\alpha/2} \cdot S/\sqrt{n}) = 1-\alpha$$

因此，常態母體、變異數未知時，母體平均數的$(1-\alpha)100\%$信賴區間為$\boxed{\bar{X} \pm Z_{\alpha/2}S/\sqrt{n}}$

## 母體平均數區間估計的總整理：

- 實際區間
  - ➢ 常態母體、母體變異數 $\sigma^2$ 已知：$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；母體平均數的 $(1-\alpha)100\%$ 信賴區間為 $\bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$
  - ➢ 常態母體、母體變異數 $\sigma^2$ 未知：$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$；母體平均數的 $(1-\alpha)100\%$ 信賴區間為 $\bar{X} \pm t_{n-1,\alpha/2}S/\sqrt{n}$
- 漸近區間
  - ➢ 非常態母體、母體變異數 $\sigma^2$ 已知：$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$；母體平均數的 $(1-\alpha)100\%$ 信賴區間為 $\bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$
  - ➢ 非常態母體、母體變異數 $\sigma^2$ 未知：$\frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$；母體平均數的 $(1-\alpha)100\%$ 信賴區間為 $\bar{X} \pm Z_{\alpha/2}S/\sqrt{n}$



## 母體比例的區間估計

- 樣本比例 $\hat{p}$ 的抽樣分配：假設某一母體中 $A$ 類別的比例為 $p$，現在用樣本比例 $\hat{p}$ 來估計 $p$。我們已知樣本比例 $\hat{p}$ 的平均數為 $E(\hat{p})=p$、變異數為 $V(\hat{p})=\frac{p(1-p)}{n}=\frac{pq}{n}$。根據中央極限定理，當樣本個數 $n\to\infty$ 時，標準化的樣本比例應趨近標準常態分配

  $$\frac{\hat{p}-p}{\sqrt{pq/n}} \to N(0,1)$$

  但左式的分母中 $pq$ 未知，因而需以 $\hat{p}\hat{q}$ 來加以估計，然而

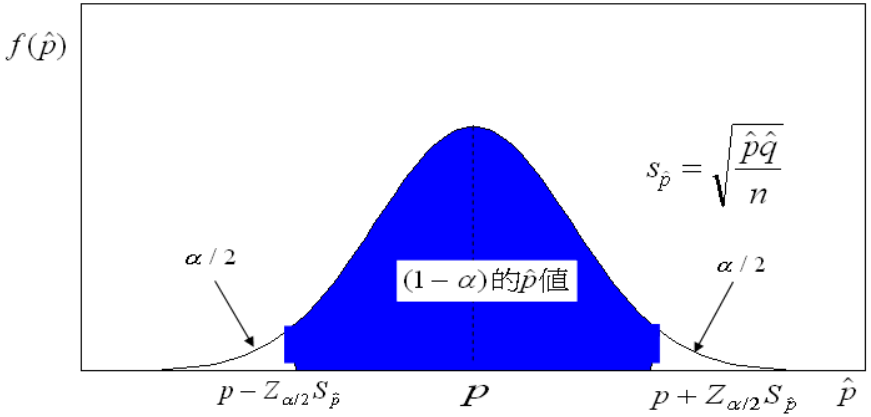
  $$\frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} = \frac{\hat{p}-p}{\sqrt{pq/n}}\sqrt{\frac{pq}{\hat{p}\hat{q}}} \xrightarrow{n\to\infty} N(0,1)$$

  因為樣本比例 $\hat{p}$ 為母體比例的一致性估計式，故當 $n\to\infty$ 時，$\hat{p}\hat{q}\to pq$，而 $\frac{\hat{p}-p}{\sqrt{pq/n}} \to N(0,1)$，因此 $\frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} \xrightarrow{n\to\infty} N(0,1)$
  - ➢ 雖說樣本比例具有實際分配，但在進行區間估計時僅漸近分配較具實用價值。



- 母體比例的區間估計：大樣本母體比例的信賴區間：給定信賴水準 $1-\alpha$，若標準常態分配比 $Z_{\alpha/2}$ 大的機率為 $\alpha/2$，則我們可得

  $$P\left(-Z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} \le Z_{\alpha/2}\right)=1-\alpha$$






因此

$$P\left(-Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}} \le \hat{p}-p \le Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\right)=1-\alpha$$

$$\Rightarrow P\left(-Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}} \le p-\hat{p} \le Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\right)=1-\alpha$$

$$\Rightarrow P\left(\hat{p}-Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}} \le p \le \hat{p}+Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\right)=1-\alpha$$

故母體比例的 $(1-\alpha)100\%$ 信賴區間為 $\hat{p}\pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$；注意，這是一個近似區間。

- 保守估計：雖然利用 $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ 來估計 $\sqrt{\frac{pq}{n}}$ 是很直覺的作法，但通常我們都以 $\sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{n}}$ 來估計 $\sqrt{\frac{pq}{n}}$。這種做法稱為保守估計，因為 $\sqrt{\frac{pq}{n}}$ 的極大值為 $\sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{n}}$；採保守估計等於是把信賴區間擴展到最大。

- **母體比例的信賴區間（保守估計）**：$\hat{p}\pm Z_{\alpha/2}\sqrt{\frac{\frac{1}{2}\frac{1}{2}}{n}}$
- 例子：某報的民意調查『馬英九的支持率 48.9%，謝長廷 27.1%』該民調『以電話號碼隨機抽樣 1119 位選民，在 95%信心水準下，抽樣誤差正負 3%』。
  - 在上述民調中，樣本比例即為 $\hat{p}=48.9\%$，若我們要建構母體比例的 95%信賴區間，則該區間為

$$\hat{p}\pm Z_{\alpha/2}\sqrt{\frac{\frac{1}{2}\frac{1}{2}}{n}}=0.489\pm1.96\sqrt{\frac{\frac{1}{2}\frac{1}{2}}{1119}}=0.489\pm0.029$$
$$=0.46\sim0.518$$

  - 『抽樣誤差正負 3%』代表的就是建構信賴區間需要加減的誤差大小($0.029\approx3\%$)。
  - 根據上述區間估計我們可以說『在 95%的信賴水準下，馬英九的支持率介於 46%~51.8%』



## 選擇樣本數以限制估計誤差

- 我們曾說過『信賴區間的長度代表的是估計誤差』、『增加樣本個數可以減少區間長度』。如果你心目之已有一個既定的最大估計誤差，要如何選定樣本數以達到這個既定目標呢？

**估計母體平均數時的樣本數**

- 以樣本平均數 $\bar{X}$ 來估計母體平均數 $\mu$ 時，估計誤差為 $|\bar{X}-\mu|$，若在 $(1-\alpha)100\%$ 信賴水準下，估計誤差不超過 $d$，則樣本數至少應該為多少？
- **常態母體、母體變異數 $\sigma^2$ 已知**：由於

$$P(-Z_{\alpha/2}\cdot\sigma/\sqrt{n}\le\bar{X}-\mu\le Z_{\alpha/2}\cdot\sigma/\sqrt{n})=1-\alpha$$
$$\Rightarrow P(|\bar{X}-\mu|\le Z_{\alpha/2}\cdot\sigma/\sqrt{n})=1-\alpha$$

在 $(1-\alpha)100\%$ 信賴水準下，估計誤差的上限為 $Z_{\alpha/2}\cdot\sigma/\sqrt{n}$ ，若估計誤差 $|\bar{X}-\mu|\le d$ 要成立，則隱含 $Z_{\alpha/2}\cdot\sigma/\sqrt{n}\le d$ 需成立，故



$$Z_{\alpha/2}^2\cdot\sigma^2/n\le d^2\Rightarrow\boxed{n\ge\frac{Z_{\alpha/2}^2\cdot\sigma^2}{d^2}}$$

- **常態母體、母體變異數 $\sigma^2$ 未知**：由於

$$P(-t_{n-1,\alpha/2}S/\sqrt{n}\le\bar{X}-\mu\le t_{n-1,\alpha/2}S/\sqrt{n})=1-\alpha$$
$$\Rightarrow P(|\bar{X}-\mu|\le t_{n-1,\alpha/2}S/\sqrt{n})=1-\alpha$$

因此，$t_{n-1,\alpha/2}S/\sqrt{n}\le d\Rightarrow t_{n-1,\alpha/2}^2S^2/n\le d^2\Rightarrow\boxed{n\ge\frac{t_{n-1,\alpha/2}^2S^2}{d^2}}$

- 非常態母體、母體變異數 $\sigma^2$ 已知：由於

$$P(-Z_{\alpha/2}\cdot\sigma/\sqrt{n}\le\bar{X}-\mu\le Z_{\alpha/2}\cdot\sigma/\sqrt{n})=1-\alpha$$
$$\Rightarrow P(|\bar{X}-\mu|\le Z_{\alpha/2}\cdot\sigma/\sqrt{n})=1-\alpha$$

因此，$Z_{\alpha/2}\cdot\sigma/\sqrt{n}\le d\Rightarrow Z_{\alpha/2}^2\cdot\sigma^2/n\le d^2\Rightarrow\boxed{n\ge\frac{Z_{\alpha/2}^2\cdot\sigma^2}{d^2}}$



- 非常態母體、母體變異數 $\sigma^2$ 未知：由於

$$P(-Z_{\alpha/2}\cdot S/\sqrt{n}\le\bar{X}-\mu\le Z_{\alpha/2}\cdot S/\sqrt{n})=1-\alpha$$
$$\Rightarrow P(|\bar{X}-\mu|\le Z_{\alpha/2}\cdot S/\sqrt{n})=1-\alpha$$

因此，$Z_{\alpha/2}\cdot S/\sqrt{n}\le d\Rightarrow Z_{\alpha/2}^2\cdot S^2/n\le d^2\Rightarrow\boxed{n\ge\frac{Z_{\alpha/2}^2\cdot S^2}{d^2}}$

**估計母體比例時的樣本數選擇**

- 估計誤差為 $|p-\hat{p}|$，而我們已知

$$P\left(-Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\le p-\hat{p}\le Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\right)=1-\alpha\Rightarrow P\left(|p-\hat{p}|\le Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\right)=1-\alpha$$

若要求估計誤差 $|p-\hat{p}|\le d$ ，則

$$Z_{\alpha/2}\sqrt{\tfrac{\hat{p}\hat{q}}{n}}\le d\Rightarrow Z_{\alpha/2}^2\tfrac{\hat{p}\hat{q}}{n}\le d^2\Rightarrow\boxed{n\ge\frac{Z_{\alpha/2}^2\hat{p}\hat{q}}{d^2}}$$

- 若採保守估計，則將 $\hat{p}\hat{q}$ 取代成 $\frac{1}{2}\cdot\frac{1}{2}=0.25$ ，亦即 $\boxed{n\ge\frac{Z_{\alpha/2}^2\cdot0.25}{d^2}}$