

## Practice#1：樣本數的決定

### 目的：

做民意調查，對於電訪的人數如何決定？在95%的信賴水準下，如果設定誤差範圍在3%，應該電訪多少人？電訪結果實際只訪問了980人，誤差範圍應該修正為多少%？你怎麼跟別人解釋95%與3%的意義？

### 練習：

1. 民意調查一般估計的是支持率  $p$ ，即該部分佔全體的比例。不論參與調查的對象有幾個，對每一個對象而言，可以視為支持與不支持兩種。支持率  $p$  與不支持率  $q$  加起來等於1。民意調查的結果即支持率  $p$  的估計值  $\hat{p}$ ，

$$\hat{p} = \frac{\text{支持某候選人的人數}}{\text{總抽樣人數}}$$

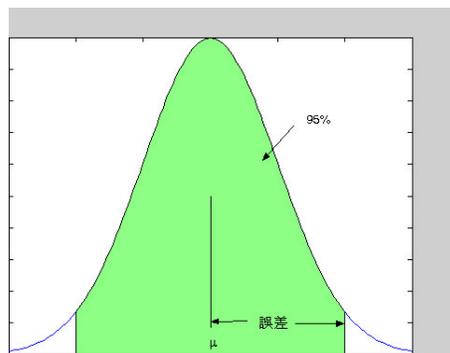
就二項分配而言，這個估計值其實就是樣本的平均數（按： $B(1, p)$ ）。當考慮估計值  $\hat{p}$  的信賴程度時，有一個問題必須先被回答： $\hat{p}$  屬於什麼分配？當然這又一次的牽涉到中央極限定理。

2. 當樣本數大時， $\hat{p}$  符合中央極限定理所描述的接近常態分配，變異數為

$$\hat{\sigma} = \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \text{where } \hat{q} = 1 - \hat{p}$$

此時，對信賴區間的計算與與前一個單元相同。你最好自行推導出上述的公式，它能幫助你瞭解一些事情的真相，特別是二項分配與常態分配間的關係。

3. 對台灣20歲以上民眾做一個民調，訪問1000人，其中有637人贊成核電廠興建。從這兩個數據能不能推論，全台灣20歲以上民眾對興建核電廠的支持度？(e.g. 推論的表達方式：在95%的信賴水準下，全台灣20歲以上民眾贊成核電廠興建的比率在60.7%到66.7%之間)
4. 上述的推論有個誤差範圍，大約是正負3%(如何得到?)。統計上的推論通常不是提供一個點的估計  $\hat{p}$  (=63.7%)，而是一個範圍(區間)的估計(60.7%~66.7%)，這個精神也是前一個練習介紹的重點。這個3%的誤差範圍與哪些因素有關呢？換句話說，就這個問題而言，如果想降低誤差範圍到1.5%(範圍愈小，可信度愈高)，應該從哪方面著手？提高抽樣的人數可行嗎？
5. 回到更實際的問題：到底要電訪多少人才能夠達到所要求的『準確度』？比較精準的描述『準確度』，可以舉例來說：在95%的信賴水準下，誤差範圍不超過



<sup>1</sup> 本單元所討論之誤差稱為隨機抽樣誤差：Random Sampling Error。是樣本統計量與母體參數之間的差距。無關乎實際抽樣動作所造成的誤差。

3%。根據這兩個條件下，該訪問多少人？計算的過程中，沒有如上題可以先算出  $\hat{p}$ ，該怎麼辦？可以假設  $\hat{p}=0.5$  嗎？ $\hat{p}$  的假設有any依據嗎？有任何可以藉助的資料提供更準確的假設嗎？

**觀察：**

1. 相關的計算公式並沒有這裡出現，同學必須自行推導或到課本去找。
2. 上面第五個練習中，計算抽樣人數時需要先知道  $\hat{p}$ 。但有時候對  $\hat{p}$  的估計所需的資料或資訊並不明顯，需要帶一點常識或採取保守的估計。因此對最後估算的樣本數也要連帶做一些解釋。
3. 看起來抽樣誤差或信賴區間是乎與母體本身的大小無關。不管是「萬」中取一還是「億」中取一，只要樣本數一樣，結果都是一樣的。這樣講對嗎？

**作業：**

1. 寫一隻程式，對於民意調查見到的三個數據：信賴水準、抽樣誤差及電訪人數，做一個『已知其二，求第三個值』。也就是：給定『信賴水準』及『抽樣誤差』，求電訪人數。或已知『信賴水準』及『電訪人數』，求『抽樣誤差』。

**參考書目：**

J.T. McClave, F.H. Dietrich, T. Sincich , 『Statistics , 7 edition』

