選修數學 進階講義 上

信賴區間與信心水準的解讀

成功高中·陳冠宏老師

信望愛文教基金會

15-3-5　信賴區間與信心水準的解讀

定理敘述

1. 點估計與區間估計

若 $x_1, x_2, \ldots, x_n$ 為來自某母體之隨機抽樣，則我們常以樣本平均數 $\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ 來估計母體平均數 μ，稱 $\overline{X}$ 為母體平均數 μ 之一點估計值。但點估計命中目標的機會是很低的，因少數樣本觀察值得到的結果要和母體平均數吻合幾乎不可能，因此我們改採用「區間估計」的方式。

即依樣本平均數對母體平均估計一個上下限的區間，並指出該區間包含母體平均的可靠度。這個區間，我們稱之為「信賴區間」，而可靠度，我們稱之為「信心水準」。

2. 民意調查

若進行一次抽樣得樣本對候選人或議題之贊成比例 $\hat{p}$ ，欲估計母體贊成比例 p，則依常態分布之經驗法則可知：

p 有 68%的機率落在區間$[\hat{p} - \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}]$內

p 有 95%的機率落在區間$[\hat{p} - 2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}]$內

p 有 99.7%的機率落在區間$[\hat{p} - 3\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 3\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}]$內，其中 n 為抽樣之樣本數

3. 常見名詞解釋

(1)抽樣誤差：母體平均與抽樣平均之差，即$|\hat{p} - p|$

(2)最大誤差：抽樣誤差之最大可能值，即若母體平均之信賴區間為$[\hat{p} - e, \hat{p} + e]$，則 e 稱為最大誤差。

68%信心水準下之最大誤差$e = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

95%信心水準下之最大誤差$e = 2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

99.7%信心水準下之最大誤差$e = 3\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

4. 保守估計

在樣本數越大的情況下，最大誤差會越小，但樣本數越大，所需調查成本會越大。因此在沒有 $\hat{p}$ 的情況下，希望知道至少需多少個樣本數，才能將最大誤差控制在 e 之內，則可取 $\hat{p}=\frac{1}{2}$，因此時 $\sqrt{\hat{p}(1-\hat{p})}$ 會有最大值。

例如在 95%的信心水準下，希望最大誤差控制在 e 之內，則保守估計樣本數 $n \ge \left(\frac{1}{e}\right)^2$

定理證明或說明

1. **信賴區間與信心水準**

若母體平均為μ，變異數為σ^2，依據中央極限定理，我們可知樣本平均$\overline{X}$會滿足一個平均為μ，變異數為$S^2=\frac{\sigma^2}{n}$之常態分布，則依常態分布之經驗法則，可知進行一次抽樣所得之樣本平均$\bar{x}$

會有 68%的機率落在$[\mu-S,\mu+S]$這個區間內

會有 95%的機率落在$[\mu-2S,\mu+2S]$這個區間內

會有 99.7%的機率落在$[\mu-3S,\mu+3S]$這個區間內

亦可視為母體平均μ

會有 68%的機率落在$[\bar{x}-S,\bar{x}+S]$這個區間內

會有 95%的機率落在$[\bar{x}-2S,\bar{x}+2S]$這個區間內

會有 99.7%的機率落在$[\bar{x}-3S,\bar{x}+3S]$這個區間內

這些區間我們稱之為「信賴區間」

而落在其中的機率我們稱之為「信心水準」

2. 候選人的支持率可視為二項分布中，成功（支持）次數的比率。若以 Y 表成功的比率，X 表參數(n,p)的二項分布中成功的次數，n 為樣本數，p 為支持的比率，則$Y=\frac{X}{n}$。

其中期望值 $E(Y)=\frac{E(X)}{n}=\frac{np}{n}=p$，

變異數 Var(Y)=$\frac{Var(X)}{n^2}=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$，標準差=$\sqrt{\frac{p(1-p)}{n}}$。

故若進行一次抽樣得樣本對候選人或議題之贊成比例 $\hat{p}$，欲估計母體贊成比例 p 則依常態分布之經驗法則可知

p 有 68%的機率落在區間$[\hat{p}-\sqrt{\frac{p(1-p)}{n}},\hat{p}+\sqrt{\frac{p(1-p)}{n}}]$內

p 有 95%的機率落在區間$[\hat{p}-2\sqrt{\frac{p(1-p)}{n}},\hat{p}+2\sqrt{\frac{p(1-p)}{n}}]$內

p 有 99.7%的機率落在區間$[\hat{p}-3\sqrt{\frac{p(1-p)}{n}},\hat{p}+3\sqrt{\frac{p(1-p)}{n}}]$內

依大數法則，當 n 夠大時，$\sqrt{\frac{p(1-p)}{n}}$ 可以 $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ 進行估計。

例如，母體平均 p 落在$[\hat{p}-\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\hat{p}+\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$的機率為 68%

即稱此區間為 68%信心水準下的信賴區間。

注意事項

信心水準的意義：m%信心水準代表依此方式進行多次抽樣得多個信賴區間，則這些信賴區間約有 m%的個數會包含到母體平均，並非代表進行一次抽樣得一個信賴區間，母體平均落在此區間的機率為 m%。

關鍵字

信賴區間，信心水準

例題 1

某校欲調查學生對「參加課外活動社團」的意願，抽樣 100 人，其中有 64 人贊成參加

(1)贊成的比例為？

(2)在 95%的信心水準下，這次調查的（最大）抽樣誤差為？

(3)在 95%的信心水準下，其信賴區間為？

Ans：(1)0.64；(2)0.096；(3)[0.544,0.736]

解：

(1)贊成的比例 $\hat{p}=\frac{64}{100}=0.64$

(2) 95%的信心水準下，抽樣誤差為

$$2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}=2\sqrt{\frac{0.64\times(1-0.64)}{100}}=0.096$$

(3) 95%的信心水準下，信賴區間為

$$[\hat{p}-2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\hat{p}+2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]=[0.64-0.096,0.64+0.096]$$

=[0.544,0.736]

例題 2

抽樣調查某大學學生住宿比率，在 95%的信心水準之下，得住宿比率的信賴區間為[0.45,0.55]
(1)此次調查共調查了多少樣本
(2)樣本約有多少人住宿

Ans：(1)400；(2)200

解：

$$[0.45,0.55]=[\hat{p}-2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\hat{p}+2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

$$\Rightarrow\begin{cases}0.45=\hat{p}-2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\\0.55=\hat{p}+2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\end{cases}\Rightarrow\begin{cases}\hat{p}=0.5\\n=400\end{cases}$$

(1)n=400；　(2)400×0.5=200

例題 3

設某新產品分別請甲、乙兩市調公司進行市場調查，已知甲、乙兩公司調查的支持比率相同，在 95%的信心水準下，甲公司所得的信賴區間長度是乙公司所得信賴區間的一半。若甲公司調查 1280 人，則乙公司約調查多少人？

Ans：320 人

解：

95%下的信賴區間為 $[\hat{p}-2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\hat{p}+2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$

區間長度為 $(\hat{p}+2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})-(\hat{p}-2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})=4\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

設甲、兩公司調查的支持度均為 $\hat{p}$ ，乙公司調查 n 人，則

$$\frac{1}{2}=\frac{4\sqrt{\dfrac{\hat{p}(1-\hat{p})}{1280}}}{4\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}\Rightarrow\frac{1}{2}=\frac{\sqrt{n}}{\sqrt{1280}}\Rightarrow n=320$$

結論：若樣本數為原來的 4 倍，則信賴區間的區間長度為原來的一半

例題 4

某媒體欲進行候選人支持度的民調，若希望在 95%的信心水準下，調查結果的誤差可以控制在正負 3%之內，則至少需調查多少個樣本？

Ans：1112

解：

95%信心水準下，最大誤差為 $2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

即求 $2\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \le 0.03$，

取 $\hat{p}=\dfrac{1}{2}$，因此時 $\sqrt{\hat{p}(1-\hat{p})}$ 會有最大值，故可得 n 最保守的估計值

$$\Rightarrow 2\sqrt{\frac{0.5\times 0.5}{n}} \le 0.03 \Rightarrow \sqrt{n} \ge \frac{1}{0.03} \Rightarrow n \ge (\frac{1}{0.03})^2 \approx 1111.11$$

故取 1112 個

例題 5

教育部想調查國中生近視的比率，在某次調查中，抽樣了 1100 位學生，其中有 605 位有近視，在 95%的信心水準下，誤差為正負 3%，試問下列敘述何者正確？

(1)根據此次抽樣所得 95%信心水準下之信賴區間為[0.52,0.58]

(2)在同樣的條件下，降低信心水準，抽樣誤差會提高

(3)若想減少抽樣誤差，可以考慮增加抽樣的人數

(4)此次調查所得信賴區間有 95%的機率包含真正近視比率

(5)若重複作 100 次的抽樣，所得的 100 個信賴區間中，大概會有 95 個包含真正的近視比率

Ans：(1)(3)(5)

解：

(1) $\dfrac{605}{1100}=0.55 \Rightarrow$ [0.95-0.03,0.95+0.03]=[0.52,0.58]

(2)抽樣誤會降低

(4)當信賴區間決定後，只有包含跟不包含。

溫故知新

習題 1

某校欲調查學生對「參加課外活動社團」的意願，抽樣 1600 人，其中有 1280 人贊成參加，則在 95%的信心水準下，其信賴區間為？

習題 2

調查學生對暑期夏令營的參加意願，在 95%的信心水準之下，得參加比例的信賴區間為[0.50,0.58]，則此次調查約調查了多少樣本？

習題 3

抽樣調查民眾對十二年國教的支持度，在 95%的信心水準之下，得支持度的信賴區間為[0.60,0.68]，則此次調查約調查了多少樣本

習題 4

某媒體欲進行候選人支持度的民調，若希望在 99.7%的信心水準下，調查結果的誤差可以控制在正負 3%之內，則至少需調查多少個樣本？

習題 5

想要了解台灣的公民對某議題支持的程度所作的抽樣調查，依性別區分，所得結果如下表：

	女性公民	男性公民
贊成此議題的比例 $\hat{p}$	0.52	0.59
$\hat{p}$ 的標準差 $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	0.02	0.04

請問從此次抽樣結果可以得到下列哪些推論？

(1)全台灣男性公民贊成此議題的比例大於女性公民贊成此議題的比例

(2)在 95%的信心水準之下，全台灣女性公民贊成此議題之比例的信賴區間為[0.48,0.56]（計算到小數點後第二位，以下四捨五入）

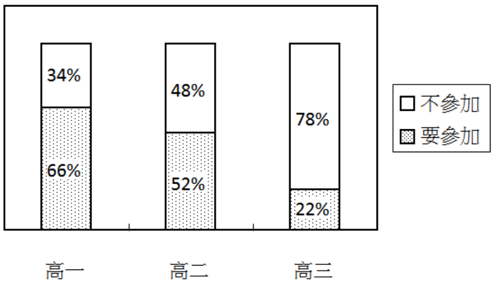
(3)此次抽樣的女性公民數少於男性公民數

(4)如果不區分性別，此次抽樣贊成此議題的比例 $\hat{p}$ 介於 0.52 與 0.59 之間

(5)如果不區分性別，此次抽樣 $\hat{p}$ 的標準差 $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ 介於 0.02 與 0.04 之間

習題 6

某縣市教育局欲瞭解高中生參加課外活動社團的意願，開學日隨機調查高一、高二、高三學生各 1067 名，詢問本學期是否要參加課外活動社團。已知該縣市的高一、高二、高三學生人數幾乎一樣多，各年級學生調查結果如右圖：



試問下列選項中的敘述，哪些是正確的？

(1)學生要參加課外活動社團之比例隨著年級增加而遞減

(2)由上述資訊可以估算全體學生要參加課外活動社團的比例

(3)在 95%信心水準下，每一個年級學生要參加課外活動社團的比例之信賴區間，都可以由題目中已知的數據算出

(4)在 95%信心水準下，三個年級的調查結果，以高一學生要參加課外活動社團的比例的信賴區間最長

(5)在 95%信心水準下，三個年級的調查結果，以高三學生要參加課外活動社團的比例的信賴區間最短

習題 7 【學測 98】

某廠商委託民調機構在甲、乙兩地調查聽過某項產品的居民佔當地居民之百分比（以下簡稱為「知名度」）。結果如下：在95%信心水準之下，該產品在甲、乙兩地的知名度之信賴區間分別為 [0.50 , 0.58]、[0.08 , 0.16]。試問下列哪些選項是正確的？

(1)甲地本次的參訪者中，54%的人聽過該產品

(2)此次民調在乙地的參訪人數少於在甲地的參訪人數

(3)此次調查結果可解讀為：甲地全體居民中有一半以上的人聽過該產品的機率大於95%

(4)若在乙地以同樣方式進行多次民調，所得知名度有95%的機會落在區間[0.08 , 0.16]

(5)經密集廣告宣傳後，在乙地再次進行民調，並增加參訪人數達原人數的四倍，則在95%信心水準之下該產品的知名度之信賴區間寬度會減半（即 0.04）

習題 8 【數乙 102】

想要了解選民對某候選人真正的支持度(支持率)p，四家媒體所做的民意調查結果如下表所示：

	媒體 A	媒體 B	媒體 C	媒體 D
$\hat{p}$	0.30	0.40	0.30	0.28
$\hat{\sigma}$	0.02	$\hat{\sigma}_B$	0.01	0.01

其中 $\hat{p}$ 表示抽樣支持度，$\hat{\sigma}=\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$，$n$ 為抽樣人數。請選出正確的選項。

(1)在 95%的信心水準之下，媒體 A 抽樣所得 p 的信賴區間為[0.28, 0.32]

(2) 如果媒體 B 抽樣的人數與媒體 A 相同，則 $\hat{\sigma}_B$ 大於 0.02

(3) 媒體 C 抽樣人數約為媒體 A 抽樣人數的兩倍

(4) 媒體 A 的抽樣支持度比媒體 B 的抽樣支持度更接近候選人真正的支持度 p

(5) 在 95%的信心水準之下，至少有一家媒體抽樣所得 p 的信賴區間會包含真正的支持度 p

解答與解析

習題 1：[0.78,0.82]

習題 2：621

習題 3：576

習題 4：2500

習題 5：(2)(4)

習題 6：(1)(2)(3)(5)

習題 7：(1)(2)

解：

(1)甲地的知名度 $p_1=\dfrac{0.5+0.58}{2}=0.54$

(2)設甲地的參訪人數為 n_1，則 $2\sqrt{\dfrac{p_1(1-p_1)}{n_1}}=0.54-0,5$

$\Rightarrow\sqrt{\dfrac{0.54\times 0.46}{n_1}}=0.02\Rightarrow n_1=621$

同理可得乙地的參訪人數 $n_2=264$

(3)只是說明抽樣的樣本中有一半以上的人，全體無法確定

(4)非信心水準的意義

(5)密集宣傳後，乙地的知名度可能改變，故未必減半

習題 8：(2)

解：

(1)[0.3-2×0.02,0.3+2×0.02]=[0.26,0.34]

(2) $\hat{\sigma}_A=\sqrt{\dfrac{0.3\times 0.7}{n_A}}<\hat{\sigma}_B=\sqrt{\dfrac{0.4\times 0.6}{n_A}}$

(3) $\hat{\sigma}_A=\sqrt{\dfrac{0.3\times 0.7}{n_A}}=0.02\Rightarrow n_A=525$

$\hat{\sigma}_C=\sqrt{\dfrac{0.3\times 0.7}{n_C}}=0.01\Rightarrow n_C=2100$，故媒體 C 抽樣人數約為媒體 A 抽樣人數的 4 倍

(4)(5)無法確定