區間估計與信賴水準

last modified August 8, 2007

許多的應用上需要估計平均值, 譬如, 汽車廠想知道某一款新車每公里的平均耗油量; 照明設備公司想知道燈泡的平均壽命; 製藥廠想知道新藥對病人血壓的平均增加量。平均值的計算很容易, 人人會算, 把試驗取得的數值加起來除以個數便是了。但是, 這樣的統計 (平均) 數值有說服力嗎? 2 個數值的平均與 200 個數值的平均是否代表不同的意涵? 這個問題必須被面對, 因為對某些試驗而言, 樣本的取得毫無困難, 而有些試驗卻只能取得少數的樣本。樣本數的多寡是否能進一步解釋其平均值?

本練習著重在進一步探討平均數 (點估計) 的內涵, 其中的「區間估計」最是常見。這個觀念的開展, 是進入統計專業的第一步, 統計人與非統計人在此分隔。就程式設計的角度來看, 如何將數學的描述與公式透過程式表達出來, 如何做「實驗」來呈現「信心水準」, 如何更精簡的組合指令, 讓程式的維護、修改更方便, 也是這個單元的重點。

本章將學到關於程式設計

變數的安排、程式的執行與維護並兼的觀念與作法。

〈本章關於 MATLAB 的指令與語法〉

指令: norminv, tinv, fprintf

1 背景介紹與練習

什麼是點估計? 什麼是區間估計? 有了點估計, 爲何還要區間估計? 兩者間有何關係? 與中央極限定理有什麼關係? 樣本數的多寡與樣本的來源 (常態或非常態母體) 對區間估計有何影響? 能不能簡單扼要的回答這些問題? 不要小看這些問題, 統計推論以此爲基礎。

範例1: 假設母體爲常態分配, 標準差爲10(如圖 1所示), 如果僅知一個樣本, 其值爲20, 對母體平均數的估計該如何表達 (點估計與區間估計各該如何表達)? 對於你的估計有多少信心?

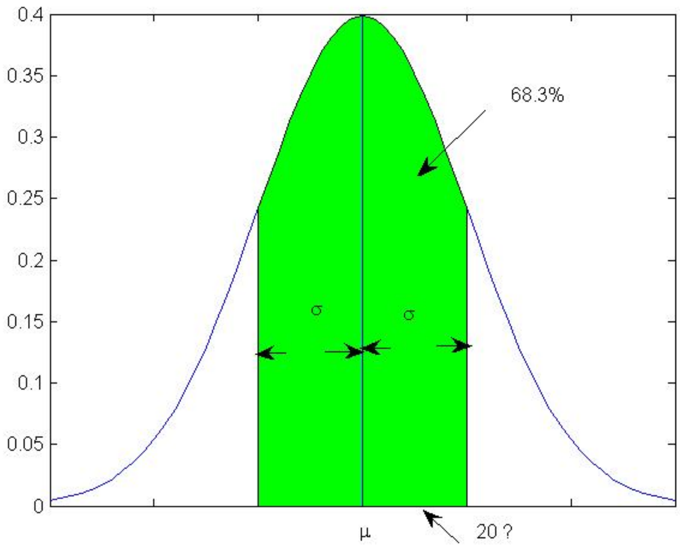

圖 1: 母體平均數μ的區間估計

圖 1爲一平均數 μ (未知), 標準差 $\sigma = 10$ 的常態分配機率密度函數圖。機率密度函數圖「敍述」了樣本可能發生的值 (範圍) 及其可能性。曲線下的面積總和爲 1, 代表所有的可能性。因此圖中陰影部分的面積可解釋爲樣本發生在 $[\mu - 10, \mu + 10]$ 範圍內的可能性, 由於剛好是一個標差的範圍, 其可能性是耳熟能詳的約 68.3% 。

當只有一個樣本時, 假設其值爲 x, 對母體平均數 μ 的點估計是別無選擇的 x, 即估計值 $\hat{\mu} = x$, 但 $\hat{\mu}$ 有多接近欲估計的母體的平均數 μ 呢?
從圖1機率密度函數的角度來看, $\hat{\mu}$ 落在 $[\mu - 10, \mu + 10]$ 之間的機率是 68.3% , 即

$$\mu - 10 \leq \hat{\mu} \leq \mu + 10 \qquad \text{with probability } 68.3\%$$

上式也可以寫成

$$\hat{\mu} - 10 \leq \mu \leq \hat{\mu} + 10 \qquad \text{with probability } 68.3\%$$

這說明了未知的母體平均數 μ 落在 $[\hat{\mu} - 10, \hat{\mu} + 10]$ 之間的可能性爲 68.3%。一般說法爲,

> 在 68.3% 的信賴水準 (Confidence Level) 下, 母體平均數 μ 的區間估計 (Confidence Interval) 爲 $[x - \sigma, x + \sigma]$。

在此將 10 換成一個 σ 來呼應 68.3%。以本範例爲例,$x = 20, \sigma = 10$, 其信賴區間的估計爲 [10 30]。

一個 σ 與 68.3% 的對應, 是從標準常態 (z 分配) 的累積機率密度函數來計算, 以往靠查表, 現在看看 MATLAB 怎麼計算 95% 的信賴水準所對應的區間範圍 (如圖2所示):

```
confidence_level=0.95;
p=(1+confidence_level)/2;
z=norminv(p,0,1)
```

MATLAB 指令 norminv 從累積機率 (第一個參數) 計算相對應的 x 值, 因此若要計算圖2的 ? 大小, 需要再加上左尾的面積, 即

$$p = \frac{1 - 0.95}{2} + 0.95 = \frac{1 + 0.95}{2}$$

這也是爲什麼在上述的 norminv 指令的第一個參數爲 (1+confidence_level)/2, 而第二與第三個參數代表標準常態的參數。這個指令計算的結果約爲 1.96, 也是個很熟

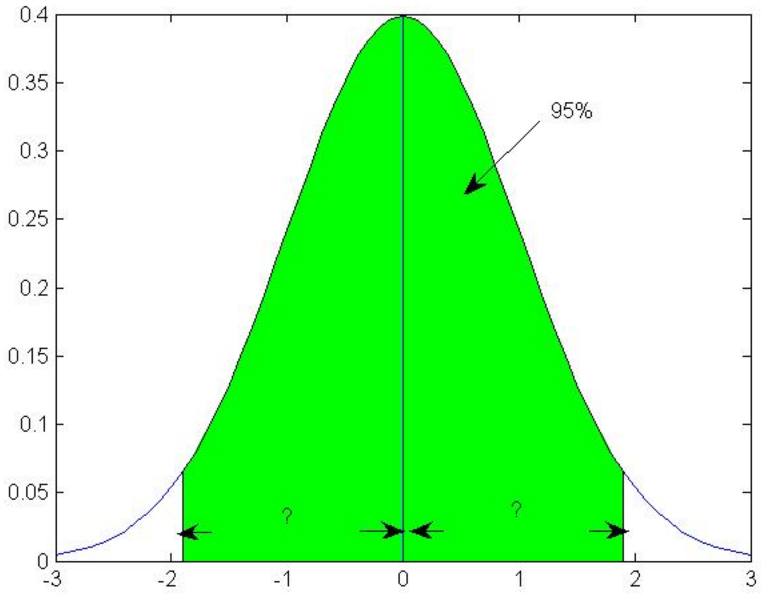

圖 2: 標準常態下的區間面積(0.95) 與範圍 (?σ) 對照圖

悉的數字。

之前的討論是基於一個樣本值 $x = 20$, 假設增加爲兩個樣本, 其值爲 (10,30) , 情況將有什麼改變呢? 注意, 我們故意讓平均值 (即對 μ 的點估計) 依然是 20, 想知道估計的品質與信心是否提高了? 如果給相同平均值的兩個樣本 (-60,100), 信心是否有些許動搖了? 在回答這些問題前, 自己先想想看, 別急著往下看。如果一點概念也沒有, 請先做做下面的實驗:

1. 兩個獨立常態變數的平均 ($x_3 = \frac{x_1+x_2}{2}$), 還是一個具常態分佈的變數嗎? 其(x_3) 均值 μ_3 與標準差 σ_3 變成什麼? 根據統計學的理論, 答案應該很淸楚, 不過此時先不管理論的結果, 我們藉著寫一支小程式來做一個小實驗, 看看結果會是什麼?

   實驗步驟: 自兩個有相同 μ 及 σ 的獨立常態母體中分別抽取一個樣本 (在執行上, 這個動作等同於從同一個常態分配的母體中抽出兩個樣本), 計算其平均值

(當作 x_3 的樣本)。重複這個看似無聊的動作很多次 (譬如,1000次), 這1000個平均數的平均值是否接近原母體的平均數 μ 呢? 那標準差呢? 對這1000個均數畫直方圖看看呈現什麼分佈? 圖3(a) 展示了這個結果。圖3(b) 則是將樣本擴大爲4個, 想觀察其平均數與變異數的差別。

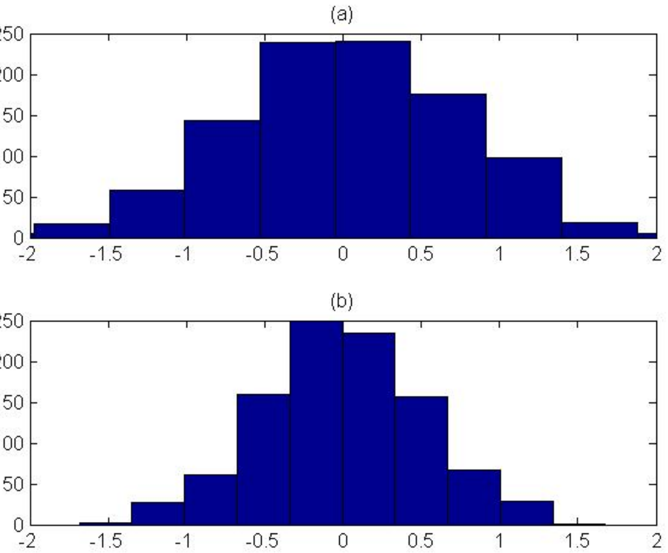

圖 3: 2個與4個變數平均值的樣本直方圖 (樣本數1000)

2. 上面這個實驗的結果, 也可以從理論上去證明 (作業1)。親自操作的實驗結果, 可以讓理論更爲鮮活, 不容易忘記, 甚至逐漸成爲下意識的常識。

範例2: 根據上面的實驗, 與範例1做比較, 當樣本數增加爲兩個, 譬如,(10,30) 與 (-60,100), 對母體平均數的點估計有何差別? 在相同的信賴水準下, 信賴區間有何改變? 這樣的改變代表的意義爲何? 這兩組樣本的區間估計是否不同?

範例 1 的實驗與理論都說明平均數的分配會隨著樣本數的增加而改變。樣本數愈多, 樣本平均數的變異愈小, 因此在相同的信賴水準下, 變異愈小的分配, 其信賴區間愈小, 估計的「品質」就愈高。圖4展示兩個變異數不同的常態分配, 在相同的面積比例 (95%) 下, 變異數小者所佔的區間較窄。

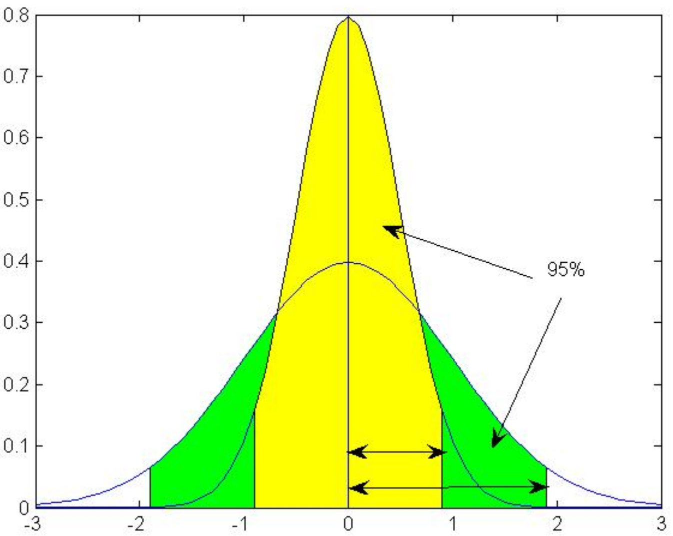

圖 4: 常態分配下的變異數大小與區間範圍

本範例中兩個樣本的平均數 $\hat{\mu} = (10 + 30)/2 = 20$, 依然服從常態分配, 但其標準差變爲 $\frac{\sigma}{\sqrt{2}}$(作業1), 在 95% 的信賴水準下, 區間估計爲

$$\hat{\mu} \pm 1.96\frac{\sigma}{\sqrt{2}} \qquad \text{即} \qquad 20 \pm 13.9$$

其中 1.96 來自 95% 的信賴水準, 分母根號裡的 2 來自樣本數。同理, 在變異數已知的情況下, 同樣是來自常態分配的兩個樣本 (-60,100), 其信賴區間的估計並不會改變。[1]

範例3: 同上, 如果有三個樣本值呢(10,20,30)? 對母體平均數的估計有何差別 (還是以 95% 的信賴水準來比較其信賴區間)。往下推論, 有100個樣本值, 其平均值已知爲20, 在 95% 的信賴水準下, 請估計母體平均數的信賴區間。

---

[1]這有點令人難以相信, 雖然平均數都「恰好」一樣, 但變異也未免太大了, 爲什麼信賴區間還是一樣呢? 其關鍵在母體變異數爲已知的假設。

以上範例主要說明樣本數與信賴區間估計的關係, 因爲樣本平均值服從常態分配, 且其變異數隨著樣本數的增加而變小, 在同樣的信賴水準下, 其母體均數的信賴區間會隨著樣本數的增加而變窄 (即估計品質變好)。

這個計算最好寫支程式, 將平均值、樣本數與信心水準都當作變動參數, 在程式執行時再輸入。寫程式前, 需將公式備妥, 確定已知條件, 如 σ 已知, 然後逐項「翻譯」成 MATLAB 指令。以本題的假設爲例, 信賴區間的公式爲

$$\hat{\mu} \pm z\frac{\sigma}{\sqrt{n}}$$

其中 σ 爲固定的標準差、z 來自信賴水準, 而 n 則是樣本數。程式如下。本範例平均值爲 20 的 3 個樣本, 在 95% 的信賴水準下, 信賴區間的估計爲 [8.684 31.316]。

```
mu=input('輸入平均值:');
n=input('輸入樣本數:');
confidence_level=input('輸入信賴水準? % :')/100;
sigma=10;
p=(1+confidence_level)/2;
z=norminv(p,0,1);
conf_intrvl=z*sigma/sqrt(n);
ci=[mu-conf_intrvl mu+conf_intrvl];
fprintf('信賴區間估計爲 %7.3f %7.3f ',ci);
```

請注意第三行要求輸入信賴水準的地方, 爲配合一般對信賴水準以百分比表示的認知, 要求輸入百分之多少的數字, 譬如輸入 95 代表 95% , 爲此在指令最後需要除以 100
[2]

範例4: 同範例1, 但是標準差未知 (這是比較合乎實際情況的假設), 對母體平均數的區間估計要怎麼做呢?

---

[2]其實在一般統計應用上, 比較常用信賴水準的另一面 α, 稱爲「誤差」, 即信賴水準爲 95% 時,$\alpha = 0.05$。因此信賴水準常寫爲 $100(1-\alpha)\%$。在程式的設計上也可以改爲以 α 爲主, 看起來專業一點。

標準差未知, 只能以估計值 s 取代, 信賴區間的估計變爲

$$\hat{\mu} \pm t\frac{s}{\sqrt{n}}$$

估計的樣本標準差 s 造成額外的誤差, 原來對應到信賴水準的 z 統計量, 必須替換爲 t 統計量, 也就是以 t 分配取代標準常態分配作爲查表的根據, 程式相對應的部分改爲

```
s=std(x)
t=tinv(p,n-1);
conf_intrvl=t*s/sqrt(n);
```

當然, 當樣本數逐漸變大時, 樣本標準差 s 會接近理論標準差 σ, 此時 t 分配也會接近標準常態分配。圖5將不同自由度的 t 分配與標準常態的機率密度函數畫在一起 (作業2), 可以清楚的看到樣本數造成的影響。

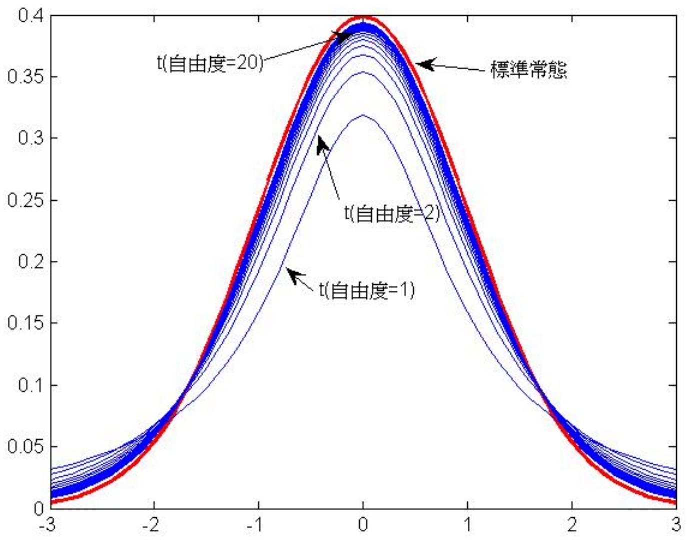

圖 5: 標準常態分配與不同自由度的 t 分配

在假設標準差未知的情況下, 再回到範例2的兩個樣本 $(10,30)$與$(-60,100)$, 在相同的信賴水準下, 其信賴區間的估計不再相同 (作業3)。

到此, 你迷路了嗎? 試著用自己的邏輯整理以上的推理, 理出個頭緒來! 如果母體不是常態分配, 樣本數大時, 可以應用中央極限定理, 其平均數依然趨近常態分配, 信賴區間的估計與前面討論的相同。但當面臨小樣本的區間估計時, 則須運用無母數統計的理論來解決。

範例5: 要畫出如圖1的陰影面積, 可以使用 MATLAB 的 area 指令, 下面的程式碼畫出圖6。

```
x=-5:0.1:5
y=normpdf(x,0,1);
plot(x,y)
hold on
area(x(31:71),y(31:71),'facecolor','green')
hold off
```

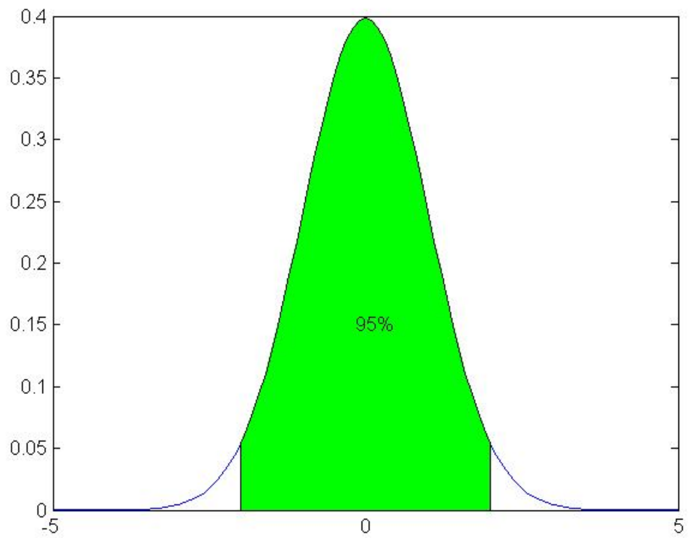

圖 6: 95% 的信賴水準示意圖

指令 area 後的選項「facecolor」決定區域的顏色。

2 觀察

1. 上面練習的數據都經過設計, 使平均值爲20。平均值雖然一樣但是因爲樣本數的不同, 導致估計品質的不同。你必須對這些不同處做出觀察, 並得出結論。
2. 對於「信賴水準」、「信賴區間」及「樣本數」間的關係是否已經徹底清楚了呢?
3. 何時該用 z 分配或 t 分配是否清楚呢? 其實 t 分配在自由度20以上, 便非常接近 z 分配 (標準常態), 因此實務上不論大小樣本均採 t 分配也是常見的。
4. MATLAB 是否提供計算信賴區間的指令呢? 試試看 ztest 與 ttest 這兩個用在假設檢定的指令。

3 作業

1. 假設兩個獨立變數 x_1, x_2, 服從常態分配 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$, 證明變數 $x_3 = ax_1 + bx_2$ 也是服從常態分配, 且其平均數與變異數分別爲

$$\mu_3 = a\mu_1 + b\mu_2, \qquad \sigma_3^2 = a^2\sigma_1^2 + b^2\sigma_2^2$$

2. 寫一支程式運用迴圈的技巧畫出圖5。
3. 重做範例2, 但假設標準差未知。
4. 同上題, 把樣本增加到7個:(5,10,15,20,25,30,35), 在 95% 的信賴水準下, 計算其信賴區間。
5. 同上題, 再把樣本增加到50個 (請到網頁下載資料4), 在 95% 的信賴水準下, 計算其信賴區間。

♣MALAB 以互動式的方式讀取外部檔案的寫法如下:

```
datafile = input('Enter data file name :', 's');
x = load(datafile);
```

程式執行時, 直接輸入完整的資料檔名稱 (必須是矩陣格式), 資料將被指定給變數 x。

6. 針對上面的練習, 整理出一支互動式程式, 在執行時讓使用者自行輸入信賴水準、資料檔 (MATLAB 指令:load)。執行結果則輸出點估計與區間估計的數據。當然結果的描述越專業越好。程式也必須判斷大樣本或小樣本, 採用適當的分配 (z 或 t)。

7. 對信賴水準 (譬如 95%) 做一個驗證。當詮釋信賴水準時, 我們說: 重複100次取樣, 在計算出來的信賴區間內, 將有95次包含母體的均數。畫一張圖, 包含100個信賴區間, 信賴區間可以用直線來表示, 如圖7所示。並且將不包含母體均數的區間用不同的顏色或線條呈現出來。觀察這個結果和你所附於的信賴水準是否符合? 程式必須夠彈性, 可以任意由使用者輸入信賴水準。當然也別忘了標準差這傢伙, 已知或未知都可以做。如果是已知, 程式必須給於使用者輸入的機會, 如果是未知, 你該知道怎麼做吧!

   ♣ 重複一百次代表程式需要用迴圈技巧。每走一圈畫一條直線, 至於線的顏色取決於該範圍是否涵蓋已知的均數。注意: 畫一條直線只需要頭尾兩個點, 信賴區間剛好提供這些數據。圖7的一百條線代表從常態分配的母體 ($N(0,1)$) 中, 每次抽出 100 個樣本, 計算其均數 $\hat{\mu}$ 之 95% 信賴水準下的信賴區間, 重複 100 次, 計算出 100 個信賴區間。虛線代表不包含母體均數的信賴區間, 共6條, 符合 95% 的意義。

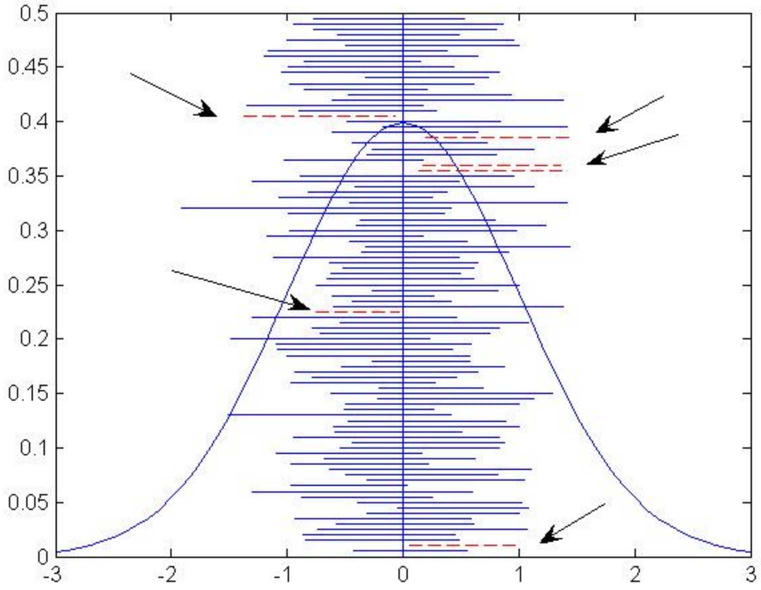

圖 7: 95%的信賴水準示意圖