信賴區間與統計檢定(測試)

取樣理論的各種分配

各種信賴區間的計算

統計假設與各種假設檢定

概述

- 如表4.1，假設有100個數量的母數，其平均值μ=26.0與變異數σ^2=17.0
- 由該母數中隨機取出10個值出來，計算其樣本平均值與樣本變異數
  - 可利用此樣本平均值與變異數來推估母數平均值與變異數
  - 但不能期望兩者會相等

表4.1 100個數之母數

13.6	22.1	24.3	23.4	30.4	24.2	20.1	23.2	28.2	25.1
31.3	30.6	31.9	24.0	23.2	29.5	29.0	19.2	28.0	24.2
28.8	24.9	26.9	30.9	19.2	27.4	26.4	30.0	21.0	16.9
26.6	17.4	22.9	21.5	24.0	27.0	20.2	31.5	26.1	24.4
31.5	34.8	25.6	27.1	21.1	25.9	24.6	17.7	21.5	26.6
27.2	26.6	28.5	26.8	27.8	25.4	27.0	25.6	27.1	31.2
25.5	31.7	22.3	30.4	26.6	21.9	29.8	30.1	22.8	33.7
29.7	32.2	20.0	22.9	26.3	24.8	28.2	19.7	26.8	31.7
24.2	22.6	31.6	27.8	26.0	27.3	15.3	27.2	29.0	26.7
26.7	31.1	23.0	26.9	31.4	25.7	27.4	27.9	23.8	27.4
				mean=	26.0			S^2=	17.0

概述

- 再由母數中隨機取樣10個，同樣地可計算出樣本平均值與變異數
  - 同樣地樣本平均值與變異數不會與母數平均值與變異數相等
  - 同時，可能不會與上一次取樣的樣本平均值與變異數相同
- 若取樣的數量逐漸增加，則樣本平均值與變異數將會逐漸地接近母數平均值與變異數

表**4.2** 增加取樣個數

取樣個數	10	20	30	40	50	60	70	80	90	100
mean=	26.5	26.9	26.5	26.4	26.3	26.2	26.0	25.9	25.9	26.0
S^2=	26.3	27.3	23.2	19.6	18.4	16.0	16.8	17.6	16.5	17.0

概述

- 表4.3為由母數中隨機取樣的組合，由表中可看出第二組的變異數最小，該組的平均值似乎是母數平均值最可靠的估值(estimate) ，但事實並非如此
- 在探討平均值與變異數時，應該要考慮樣本大小的因素
  - 如兩組之變異數一樣，則由*30*個樣本計算的平均值或許要比由*5*個樣本所計算者更為可靠
- 取樣分配理論(Sampling distribution theory)
  - 在探討樣本集合、樣本大小以及所計算的平均值與變異數等之間的關係
  - 根據此理論，平均值與變異數的估值將隨取樣而變，即隨樣本不同而改變

表4.3 由母數中隨機取樣組合

											mean=	S^2=
第1組	31.5	34.8	25.6	27.1	21.1	25.9	24.6	17.7	21.5	26.6	25.6	24.9
第2組	27.2	26.6	28.5	26.8	27.8	25.4	27.0	25.6	27.1	31.2	27.3	2.7
第3組	26.7	31.1	23.0	26.9	31.4	25.7	27.4	27.9	23.8	27.4	27.1	7.3
第4組	28.8	24.9	26.9	30.9	19.2	27.4	26.4	30.0	21.0	16.9	25.2	22.2

概述

- 推估子(estimator)是用來計算估值的函數，例如，前述第二章計算樣本平均值與樣本變異數的公式，就是用來推估母數平均值與變異數的推估子
  - 前述已證明，這些估值將隨樣本而變化，也各自擁有自己的母數分配
- 下一節將定義三種分配，並用以描述與評估平均值與變異數估值的可靠度
- 使用這些分配，我們可以預期在任一已知的信心水準下，所計算的估值有多合適(how good)
  - 換言之，在選定的機率水準下，可預期平均值與變異數會落於某特定的信賴區間(confidence interval)內

取樣理論所使用的分配

- 有三種分配被使用
  - χ^2分配
    - 比較母數變異數與樣本變異數的關係
  - *t* 分配
    - 比較母數之平均值與樣本之平均值
  - *F* 分配
    - 比較兩個樣本之變異數

取樣理論所使用的分配

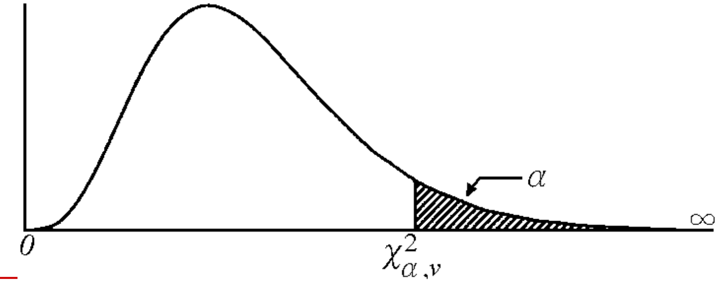

圖4.1 χ^2分配

- χ^2分配
  - 或稱卡方分配(chi-square distribution)，主要是基於樣本的多餘數υ (number of redundancies或稱自由度degree of freedom)，來比較母數變異數與樣本變異數的關係
  - 假設母數為N(μ,σ²)，由其中隨機取樣n個觀測量，則χ^2取樣分配為$\chi^2 = \upsilon S^2/\sigma^2$
  - 分配的圖形如圖4.1所示，多餘數$\upsilon=n-1$
  - 附表D.2為χ^2分配曲線表(自由度從*1*至*120*)，表中所列的是曲線下從指定的χ^2到∞的面積，即圖中斜線部分的面積，可由行的υ與列的α相交處的值查得
    - 如，求自由度$\upsilon=10$，$\alpha=0.025$之χ^2值為*20.48*
    - 因為χ^2分配為不對稱曲線，要查圖形左端的面積，必須用1-α計算
      - 如，$\upsilon=10$，左端2.5%曲線下，$\alpha=0.975=1-0.025$之χ^2值為*3.25*
  - χ^2分配可用於決定母數變異數的預期發生範圍，
    - 這是基於
      - 某個特定百分機率
      - 樣本之變異數
      - 樣本之自由度

取樣理論所使用的分配

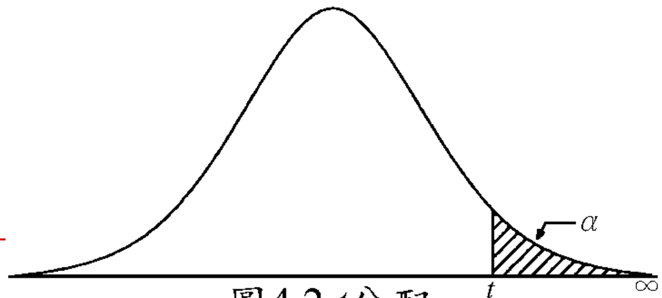

圖4.2 *t*分配

- *t (Student)*分配
  - 基於樣本多餘數υ，*t* 分配是用來比較母數平均值與樣本平均值
  - *t* 分配與常態分配非常類似(均為對稱曲線)，常態分配是用在母數，*t* 分配則用在樣本
    - 若樣本個數少於*30*時，*t* 分配又較常態分配優先採用
      - 對分析測量資料而言，*t* 分配是一個重要的分配
  - 若z∈N(0,1)，χ^2為自由度υ之卡方隨機變數，且兩者為獨立變數，則可定義*t*變數為如下的公式，並具有*t*分配
  - 附表D.3為*t*分配曲線表，如圖4.2所示，α為某 *t*值至∞之曲線下百分比面積，或圖中斜線部分的面積，表中 *t*值為基於自由度υ下，與α對應的值
    - 例如，υ=*10*，*α=0.025*之*t =2.228*；因為曲線之對稱性，故可得υ=*10*時，在-*2.228*與-∞之間，曲線下面積也是0.025
  - 根據樣本之平均值、變異數與自由度，*t*分配可用來建立母數平均值之信賴區間(見例4.1)
  - *t*分配也可用來決定樣本平均值是否足以當作母數平均值的可靠估值(見例4.6)

$$t = \frac{z}{\sqrt{\chi^2 / \nu}}$$

取樣理論所使用的分配

- *F*分配
  - 此分配是用來比較兩組樣本的變異數
  - 若χ_1^2為χ_2^2為兩個自由度分別為υ_1與υ_2之卡方隨機變數，且兩個變數為互相獨立，則定義：$F = (\chi_1^2/\upsilon_1)/(\chi_2^2/\upsilon_2)$具有*F*分配
  - 附表D.4為*F*分配曲線表，同樣表示曲線下由某點$F(\alpha,\upsilon_1,\upsilon_2)$到+∞的面積
    - 由於*F*分配有兩個自由度，因此僅能列出幾個常用的α值(*0.20, 0.10, 0.05, 0.025, 0.01, 0.005, 0.001*)
    - 例如對自由度$\upsilon_1=5$， $\upsilon_2=10$，求曲線上端(右端) *0.025*之*F*值時，應查α=*0.025*之表，由表中$\upsilon_1=5$與$\upsilon_2=10$相交處即可求得*F*值為*4.24*
  - *F*分配可用來回答：兩個樣本是否來自同一個母數？
    - 例如，兩個樣本變異數分別為：S_1^2與S_2^2，若兩個樣本變異數是代表同一個母數變異數，則其母數變異數的比值(σ_1^2/σ_2^2)應等於*1*
      - 於4.7節將討論以*F*分配來建立這個比的信賴區間
      - 於4.11節將討論以*F*分配來進行兩個變異數的比值是否等於1的檢定(測試)

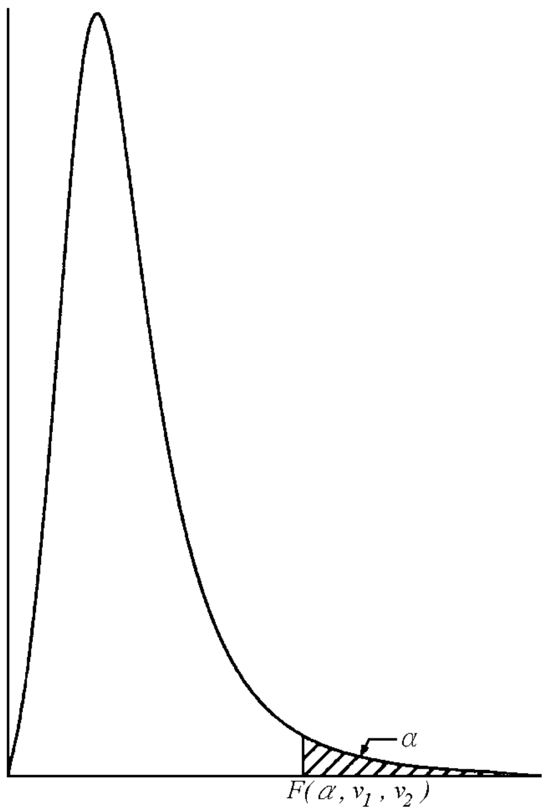

圖4.3 *F*分配

平均值的信賴區間：*t*統計量

- 根據樣本平均值與標準差，N(0,1)可用來預測母數平均值的範圍
  - 值得注意的是，$N(\mu,\sigma^2)$是基於整個母數，且在樣本數量少的情況下，可預見它會有所變化，也因此，才發展出*t*分配
    - 若查表D.3，可發現當樣本數量是無限大時，所查出來的*t*值，與從常態分配所查出來的值是相同的
      - 通常樣本數量超過*30*時，表3.2的數值即可用來建立母數平均值之區間，但若樣本數量少於*30*，仍應使用*t*分配之*t*值來建構母數平均值之信賴區間

平均值的信賴區間：*t*統計量

- 信賴區間公式的推導
  - 由N(μ,σ²)的母數中取樣，並計算樣本平均值 $\bar{y}$
  - 將N(μ,σ²)進行標準化，令 $z=(\bar{y}-\mu)/(\sigma/\sqrt{n})$
  - 將z與χ²代入*t*分配的定義式中，經過簡化後即可獲得*t*統計量
  - 因此，要計算母數平均值的信賴區間，必須先決定(1-α)的面積大小

$$t=\frac{z}{\sqrt{\chi^2/v}}=\frac{(\bar{y}-\mu)(\sigma/\sqrt{n})}{\sqrt{(vS^2/\sigma^2)/v}}=\frac{(\bar{y}-\mu)(\sigma/\sqrt{n})}{S/\sigma}=\frac{\bar{y}-\mu}{S/\sqrt{n}} \qquad (4.4)$$

平均值的信賴區間：t統計量

- 例如，對一個95％之信賴區間(圖4.4中非斜線部分)，可將0.95之百分比點置於t分配之中央部分，因此兩端各留0.025，如圖4.4中之斜線部分
- 由表D.3可查得t分配兩端各$\alpha/2$的t值。對平均值為$\bar{y}$、變異數為S^2之樣本，表示這個區域面積之機率為

$$P(|z|<t)=1-\alpha \qquad (a)$$

$$P(\bar{y}-t\frac{S}{\sqrt{n}}<\mu<\bar{y}+t\frac{S}{\sqrt{n}})=1-\alpha \qquad (4.5)$$

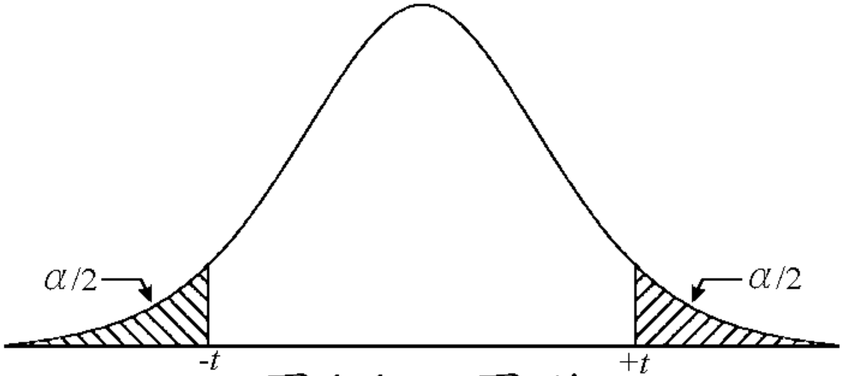

圖4.4 $t_{\alpha/2}$圖形

平均值的信賴區間：*t*統計量

- 因此，若已知樣本平均值、$t_{\alpha/2,\upsilon}$、*n*與樣本標準差時，則母數平均值的$(1-\alpha)$或是誤差區間為(4.6)式所示
- *例4.1* 控制測量時，對某個方向觀測*16*次，秒位之平均值為*25.4″*，標準差=*±1.3″*，試求(1)母數平均值之*95%*信賴區間；(2)所求區間再與由標準常態分配所求*t*值而決定之區間作一比較
  - 本例的信心水準為0.95；因此，α=0.05，得α/2=0.025
  - 信賴區間是以μ為中心，兩端為由附表D.3所查得的± $t_{\alpha/2,\upsilon}$
  - 因為n=16，故υ=16-1=15；由附表D.3可查得$t_{\alpha/2,\upsilon}$ =2.131
  - 由式(4.6)知
  - 24.7=25.4-2.131(1.3/√16)<μ<25.4+2.131(1.3 /√16)=26.1
  - 若樣本數量大，由常態分配可查得*95%*信心水準的t值為1.960，而平均值的標準誤差為±1.3/√16=±0.325
  - 因此，母數平均值的範圍為25.4 ±1.960×0.325或(24.8, 26.0)
  - 由於樣本數量小，t分配提供了母數平均值比較大的範圍，若樣本數量大到無限大，則t分配所查得的t值為1.960，與常態分配所查得的相同

$$\bar{y} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{y} + t_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (4.6)$$

信賴區間有效性的測試

- 信賴區間理論有效性的測試
  - 以電腦常態亂數產生器產生1000組樣本，每組包含16個樣本
    - 假設這些樣本是從平均值為25.4、標準差為±1.3的母數中所隨機取樣的
    - 由每組樣本的平均值與標準差，以95%的信心水準來推估母數的平均值，再與母數的真實平均值比較
      - 若理論是有效的，則所推估的區間應該包含母數平均值
    - 這些結果列於附表E.1
      - 區間未包含母數平均值的，在表中以*表示，在1000組中共有50組有*號，即有5%有*號；換言之，有95%的樣本組包含母數平均值
      - 這證明利用式(4.6)來計算母數平均值的界限是可行的

樣本大小的選擇

- 在測量實務中，常遭遇的共同問題是，需要決定重複觀測的次數，以滿足量測精度的要求
  - 實際上，S的大小並無法絕對控制，再者，由式(4.6)可看出，信賴區間僅能藉改變重複觀測的次數來控制
    - 通常，有較大的樣本數量，其信賴區間會較小
  - 根據所選擇的信心水準α，母數平均值μ之範圍為式(b)所示
  - 令I為母數平均值μ座落區間的一半，則得式(4.7)，再重新整理可得式(4.8)

$$\bar{y} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (b)$$

$$I = t_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (4.7)$$

$$n = \left(\frac{t_{\alpha/2}S}{I}\right)^2 \qquad (4.8)$$

樣本大小的選擇

- 實際應用式*(4.8)* 時，因為資料尚未收集，故$t_{\alpha/2}$與S均為未知；此外，量測個數與多餘個數為原本要解決的問題，也是未知
- 因此，式*(4.8)*必須加以修正成，利用標準常態隨機變數z，而其t值必須與υ或n無關，即修正成式(4.9)
- *例***4.2** 由水平控制網的前級分析，知道在95%信心水準下，所有角度量測都必須在$\pm 2''$之內；如果單一角度量測之標準差在$\pm 2.6''$之內；則需重複量測的次數為若干？
  - 根據*(4.9)*式，$n=(1.960\times 2.6/2)^2=6.49$；最接近之偶數(由於儀器需要正倒鏡觀測，以消除儀器系統誤差)個數為*8*，故取*8*次

$$n=\left(\frac{t_{\alpha/2}\sigma}{I}\right)^2 \qquad (4.9)$$

母數變異數的信賴區間

- 由式*(4.1)* $\chi^2 = \upsilon S^2/\sigma^2$，知道母數變異數$\sigma^2$的信賴區間是要根據$\chi^2$統計量來計算
  - χ^2雙尾端之百分點面積已列於表D.2，表列值為自χ^2至∞之曲線下百分比面積（以$\chi_\alpha{}^2$表之）
    - 即對一已知多餘數υ而言，$P(\chi^2 > \chi_\alpha{}^2) = \alpha$
  - 因為χ^2分配為非對稱曲線，所以要求得χ^2分配低尾端的機率，必須求得高尾端的機率$\chi^2{}_{1-\alpha}$，且使
    - $P(\chi^2 > \chi_{1-\alpha}{}^2) = 1-\alpha$
  - 故對χ^2的機率如式(4.10)所示
  - 將式(4.1)代入得式(4.11)
  - 因此， 在(1-α)100%信心水準下，母數變異數的信賴區間為式(4.13)

$$P(\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}) = 1-\alpha \qquad (410)$$

$$P(\chi^2_{1-\alpha/2} < \frac{vS^2}{\sigma^2} < \chi^2_{\alpha/2}) = P(\frac{\chi^2_{1-\alpha/2}}{vS^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{\alpha/2}}{vS^2}) = 1-\alpha \qquad (411)$$

$$\frac{vS^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{vS^2}{\chi^2_{1-\alpha/2}} \qquad (413)$$

母數變異數的信賴區間

- *例4.3* 某觀測者以1″讀儀器照準一遠方明確的目標並讀數，共 20次，其樣本標準差為±1.8″，試求95％信心水準下，σ^2 的信賴區間為何？
  - 因$1-\alpha=0.95$，或$\alpha=0.05$，$\alpha/2=0.025$，$\upsilon=20-1=19$；查表D.2得： $\chi^2_{0.975}=8.91$
  - $\chi^2_{0.025}=32.85$
  - 由式(4.13)得

$$\frac{(20-1)1.8^2}{32.85}<\sigma^2<\frac{(20-1)1.8^2}{8.91}$$ 或$1.89<\sigma^2<6.91$

95%信心水準下σ^2的信賴區間

兩母數變異數比的信賴區間

- 另一常被用的統計程序是比較兩母數變異數的比
  - 若兩樣本集合是從常態母數中隨機取樣，則可求得σ_1^2/σ_2^2比的取樣分配
  - σ_1^2/σ_2^2的信賴區間是基於F分配，將式(4.1)代入式(4.3)得式(4.14)
  - 為求σ_1^2/σ_2^2的信賴區間，需求解F分配的高低尾端值
    - 表D.4之F分配表並未列出低尾端部分值，實際上也不需要，因為 $F_l = F_{1-\alpha/2,\upsilon_1,\upsilon_2} = 1/F_{\alpha/2,\upsilon_2,\upsilon_1}$

左尾端值

    - 求得F的信賴區間的機率
      - $P(F_{1-\alpha/2,\upsilon_1,\upsilon_2} < F < F_{\alpha/2,\upsilon_2,\upsilon_1}) = 1-\alpha$
    - 重新排列後，得式(4.16)

$$F = \frac{v_1 S_1^2/(\sigma_1^2/v_1)}{v_2 S_2^2/(\sigma_2^2/v_2)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \qquad (414)$$

$$P(F_l < F < F_u) = P(F_l \frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} F_u)$$

$$= P(\frac{1}{F_u} \times \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \times \frac{1}{F_l}) = 1-\alpha \qquad (4.16)$$

兩母數變異數比的信賴區間

- 將F的左右尾端值代入式(4.16)，即可得到σ_1^2/σ_2^2比的(1- α)信賴區間，式(4.18)
  - 注意式*(4.18)* 中，上下限之自由度恰好相反
- 式(4.18)常被用來進行平面控制的平差與分析
  - 在平差時，控制點會固定點位的位置與方位；當觀測資料要與大於最小約制(minimally constrained)的控制點吻合時，若控制點的坐標不一致(即反算後無法與距離以及方位角一致)，則觀測資料將被平差去符合控制點，而導致資料扭曲
    - 導線測量的最小約制為一個控制點的坐標與一個邊的方向
  - 爲了避免此情況發生，必須找出不一致的控制點，其做法是
    - 先以最小約制進行平差，得一樣本參考變異數S_1^2
    - 加入所有控制點再進行平差，又得另一樣本參考變異數S_2^2
    - 比較兩個變異數，若沒有控制點不一致的情形，則其比值應該等於1

$$\frac{S_1^2}{S_2^2}\times\frac{1}{F_{\alpha/2,v_1,v_2}}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\times F_{\alpha/2,v_2,v_1} \qquad (4.18)$$

兩母數變異數比的信賴區間

- *例4.4* 假設最小約制三邊網平差之自由度=24，其參考變異數為0.49，全約制網平差之自由度=30，參考變異數為2.25，兩變異數比，在95％信心水準下的信賴區間為何？此區間是否包含1這個值？是否有理由說控制點的值不一致？
  - 依(4.18)式，分子部份對應全約制網平差，其變異數為2.25，自由度 $\upsilon_1=30$；而分母部份對應最小約制網平差，其變異數為0.49，自由度 $\upsilon_2=24$；因 $\alpha=0.05$，查表D.4得 $F_{0.025,\upsilon1,\upsilon2}$ 與 $F_{0.025,\upsilon2,\upsilon1}$，故變異數比在95％之信賴區間為 $(2.25/0.49)(1/2.21)<\sigma_1^2/\sigma_2^2<(2.25/0.49)\times 2.14$ 或 $2.08<\sigma_1^2/\sigma_2^2<9.83$
  - 區間並不含1，或 $\sigma_1^2/\sigma_2^2 \neq 1$，或稱至少在95％的確定水準下，$\sigma_1^2 \neq \sigma_2^2$
  - 由於變異數的大小與誤差有關，因此可以說在全約制平差中，觀測與控制之間出現不符情形，原因可能是控制點的坐標不一致，或者是觀測量有未改正的系統誤差

假設檢定

- 在統計學上，區間的實際數值其實並不是最重要，更重要的是如何回答問題
  - 例如，從母數中所計算的統計量是否與預期的一致？
  - 在例4.4中，『此信賴區間是否包含了預期的比值』就要比『兩變異數比信賴區間的實際界限』來的重要
  - 用來檢定統計量(或統計子)有效性的程序，稱之為假設檢定(hypothesis testing)

假設檢定

- 假設檢定的基本元素有四
  - 零假設(null hypothesis)
    - H_o，是比較樣本統計量與母數統計量的敘述
      - 隱含樣本統計量是「預期的」母數的一部分
      - 如，在例4.4中，這就是指變異數比在統計上相當於1
  - 對立假設(alternative hypothesis)
    - H_a，在接受要拒絕零假設時，就表示樣本統計量是從另一個母數中決定的(不是從「預期的」母數中決定的)
      - 在例4.4中，對立假設為變異數比不等於1

假設檢定

- 檢定統計量(testing statistic)
  - 是從樣本資料中計算得來的，其數值是用來決定是否要拒絕零假設的依據
    - 這表示，樣本統計量的計算並非從「預期的」母數中得來的
    - 在例4.4中，當變異數比在統計上不等於1的時候，就拒絕零假設
- 拒絕區域(rejection region)
  - 在零假設被拒絕時，檢定統計量的數值
    - 參照信賴區間時，這個數值就是信賴區間界限發生之處
    - 當計算的檢定統計量比定義的拒絕區域的數值大的時候，就相當於零假設的樣本統計量已經超出了信賴區間的界限

假設檢定

- 統計檢定可能會有錯
  - 回顧例4.4，是在建立一個95%信賴區間，此區間仍有5%的機會是錯的
    - 即，比預期的變異數比值大的數值，仍可能與觀測量母數是一致的
  - 這就是為何需要進一步實施統計檢定分析的理由
- 對統計量做決定時，會有兩種基本的誤差產生
  - 型1誤差(Type I error)：有效的統計量被拒絕(以符號α表示)
    - 當事實為真時，零假設被拒絕
  - 型2誤差(Type II error)：無效的統計量被接受(以符號β表示)
    - 當事實為假時，零假設未被拒絕
  - 這兩種誤差並非來自相同的母數，因此，犯這兩種誤差的機率並不直接相關

假設檢定

- 在做決策時必須要先判斷何種誤差對實情影響較大，再者，該決策必須是基於犯這兩種誤差所產生的結果
  - 例如，合約上規定95%的點位精度要在±1公分以內，為了確保能滿足合約規範，測量員會傾向犯型1誤差
  - 同一個測量員為了支援較小比例尺測圖的控制精度±3公分，他可能會傾向犯型2誤差
  - 此兩種案例中，都必須計算所犯型1與型2誤差的機率，用以評估統計檢定推論的可靠度

假設檢定

- 表4.4為決策、α與β機率以及接受或拒絕零假設的關係
- 在圖4.5中，左邊的分配曲線為H_0的資料圖形
  - 表示H_0為真的曲線
- 圖4.5中，右邊的分配曲線表示H_a為真的圖形
- 這兩個分配曲線可視為觀測量只含偶然誤差(左邊分配)vs.觀測量含有錯誤(右邊分配)
  - 圖中可看出位於左邊曲線α區域內的有效觀測量，在α顯著水準下會被拒絕

表4.4 統計檢定的關係表

狀況	決策	
	接受H_0	拒絕H_0
H_0為真	正確決策：P=1-α	型1誤差：P=α
H_0為假	型2誤差：P=β	正確決策：P=1-β

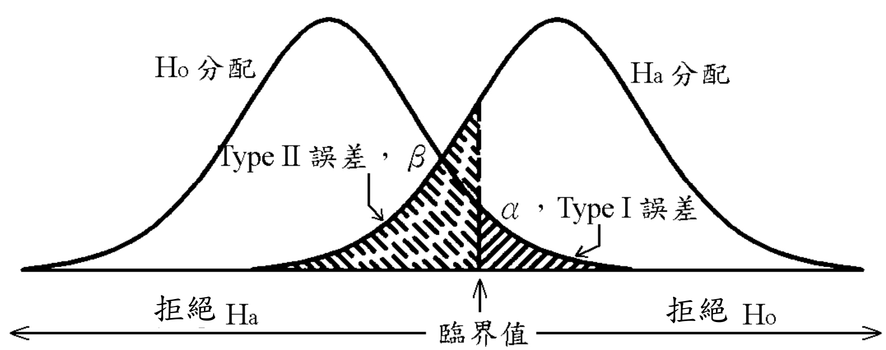

圖4.5 Type I與II 誤差的圖示

假設檢定

- α表示犯型1誤差的機率，稱之為檢定的顯著水準(significance level of the test)
- 再者，位於右邊曲線β區域內的資料，在β顯著水準下會被接受。對一個為真的對立假設來說，其檢定能力(Power of the test)為1-β
- 計算β或1-β的方法並不明確，甚至相當困難，因為對立假設的分配並不知道
- 在統計檢定中，是以資料不支持統計量是來自零假設分配，來證明對立假設為真
  - 就是以拒絕零假設來證明對立假設的成立
  - 因此，只能得到型1誤差，而其做出錯誤決定的機率則為α

假設檢定

- *例4.5* 設有一*10000*人的母數，流行病毒的檢定有*95%*信心水準，因此顯著水準α為*0.05*，假設*9200*人流行病毒測試為陰性，*800*人測試為陽性。
  - 檢定為陽性的*800*人當中，有*5%(40*人*)*檢定不正確(假陽性)，即檢定為陽性，但卻無病毒
  - 這就是在α顯著水準下，犯了型1誤差
  - 相同地，$5\% \times 9200 = 460$人檢定為陰性，而實際卻有病毒(假陰性)
  - 這就是在β $(=460/10000=0.046)$機率下，犯了型2誤差
  - 再者，本例的檢定能力$1-\beta=0.954$

假設檢定

- 在已知H_0的情況下，要確定犯了型1誤差的機率是有可能的；但在固定的α水準與樣本大小n下，犯型2誤差的機率可能不知道
  - 若H_0與α固定，則要增高檢定能力只能藉由增加樣本大小來達成
  - 由於檢定能力可能低或不知道，因此，統計學家都說檢定不拒絕(fail to reject) H_o，而不是做任何接受的敘述
  - 相同的狀況會發生在測量觀測量
    - 若距離觀測量含有大的系統誤差，則可藉由全約制平差來偵測這些誤差，因此，拒絕H_0
    - 但若距離觀測量含很小的系統誤差，則偵測出誤差的能力會低
      - 當某些信賴區間發生在H_0的拒絕區域內，不能說H_o應該被接受，因為無法決定那些偵測不出之小系統誤差的機率

假設檢定

- 假設檢定的基本步驟
  1. 根據問題的要求建立零假設與對立假設
  2. 選取一個合適的統計量，該統計量的分配必須是已知的
  3. 指定顯著水準α的值，由此確定其臨界值(區間界限)
  4. 利用樣本觀測值計算統計量的數值，看它是落於拒絕區域還是接受區域，從而決定是否拒絕H_0

母數平均值的假設檢定

- 檢定樣本平均值是否等於一已知值時，H_0有兩種方式：單尾或雙尾檢定
  - 單尾檢定主要是決定樣本平均值是否大於或小於母數平均值
  - 雙尾檢定主要是樣本平均值是否不等於母數平均值

	單尾檢定	雙尾檢定
	$H_o: \mu = \overline{y}$	$H_o: \mu = \overline{y}$
	$H_a: \mu > \overline{y} (\mu < \overline{y})$	$H_a: \mu \neq \overline{y}$
檢定統計量	$t = (\overline{y} - \mu)/(S/\sqrt{n})$	(4.19)
拒絕H_0的臨界值	$t > t_\alpha (or \ \ t < t_\alpha)$	$\lvert t \rvert > t_{\alpha/2}$

母數平均值的假設檢定

- *例4.6* 某基線率定長度為*400.008m*，以EDM重複量測*20*次後，其平均值為*400.012m*，標準差為$\pm 0.002m$；在*0.05*顯著水準下，量測距離是否與率定距離有顯著之不同？
  - 這是一個平均值的雙尾檢定問題
    - $H_o：\mu = 400.012$
    - $H_a：\mu \neq 400.012$
  - 計算t統計量
  - 拒絕區域為$t=8.944>t_{\alpha/2}=2.093$　因此，拒絕H_0

在0.05顯著水準下，量測距離的平均值與率定距離有顯著之不同，意味著至少有5%決定是錯的

$$t = (\bar{y} - \mu)/(S/\sqrt{n}) = (400.012 - 400.008)/(0.002/\sqrt{20}) = 8.944$$

$$400.011 \approx 400.012\text{-}2.093 \times \frac{0.002}{\sqrt{20}} \leq \mu \leq 400.012 + 2.093 \times \frac{0.002}{\sqrt{20}} \approx 400.013$$

率定長度400.008並沒有落於95%的信賴區間

母數變異數σ^2的假設檢定

- 除量測距離與率定距離的檢定外，測量員還想知道儀器的精密度是否如公佈的精密度
  - 精度度與標準差直接相關，即與變異數相關
  - 公佈的精密度可視為母數變異數的開方，而儀器的精密度則為樣本變異數的開方
- 要比較這兩者，可使用χ^2分配，即使用χ^2統計量來檢定
  - 同樣地，檢定方式有雙尾與單尾兩種

	單尾檢定	雙尾檢定
	$H_o: S^2 = \sigma^2$	$H_o: S^2 = \sigma^2$
	$H_a: S^2 > \sigma^2$ (or $H_a: S^2 < \sigma^2$)	$H_a: S^2 \neq \sigma^2$
檢定統計量 →	$\chi^2 = \nu S^2 / \sigma^2$	(4.20)
拒絕H_0的臨界值	$\chi^2 > \chi^2_\alpha (or\, \chi^2 < \chi^2_{1-\alpha})$	$\chi^2 < \chi^2_{1-\alpha/2} (or\, \chi^2 > \chi^2_{\alpha/2})$

母數變異數σ^2的假設檢定

- 拒絕區域由式(4.11)決定，若以圖形表示則如圖4.6所示
- *例4.7* 測量公司經理希望所有測量員在使用一部特殊儀器時，讀數能在$\pm 1.5''$之內，為了檢定這個數值，先檢定某資深工程師，其結果為：讀數30次， $S_r = \pm 0.9''$，這是否證實在5%顯著水準下，在$1.5''$之限度內？
  - 本例是希望讀數在1.5"之內，即希望變異數要小於等於1.5^2，也就是說是母數變異數的單尾假設檢定
    - H_o：$S^2=\sigma^2$
    - H_a：$S^2>\sigma^2$
    - 檢定統計量$\chi^2=(30-1)\ (0.9)2/1.52=10.44$
    - 查表D.2，得$\chi^2_{0.05,29}=42.56$
    - 因為$10.44\chi^2<\chi^2_{0.05,29}=42.56$，故不拒絕$H_0$

不拒絕H_o並不意謂$\pm 1.5''$是有效的

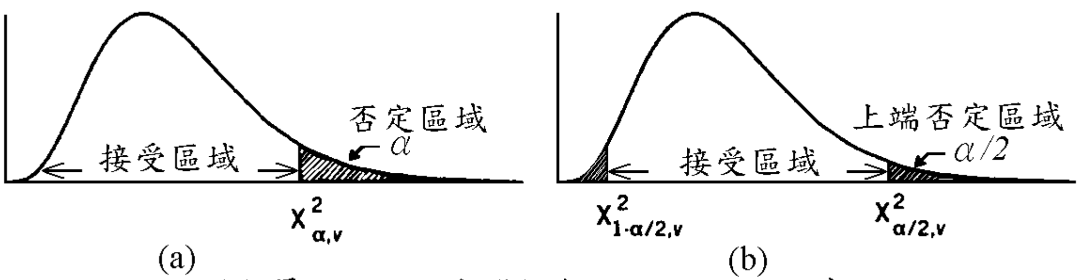

圖4.6 (a)單尾檢定與(b)雙尾檢定之示意圖

兩母數變異數比的假設檢定

- 控制資料同樣含有誤差
- 就如同在例4.4中，為了偵測控制點的誤差，必須進行最小約制與全約制的最小自乘平差
  - 這兩種平差完成後，比較平差後的參考變異數
  - 若控制沒有誤差且觀測資料沒有系統誤差，則兩者的變異數比值應接近1
  - 因此，是一種兩母數變異數比的假設檢定

單尾檢定

$$H_o: \frac{S_1^2}{S_2^2} = 1 (即 S_1^2 = S_2^2)$$

$$H_a: \frac{S_1^2}{S_2^2} > 1 (即 S_1^2 > S_2^2)$$

$$\left(H_a: \frac{S_1^2}{S_2^2} < 1 (即 S_1^2 < S_2^2) \right)$$

雙尾檢定

$$H_o: \frac{S_1^2}{S_2^2} = 1 (即 S_1^2 = S_2^2)$$

$$H_a: \frac{S_1^2}{S_2^2} \neq 1 (即 S_1^2 \neq S_2^2)$$

兩母數變異數比的假設檢定

	單尾檢定	雙尾檢定
檢定統計量	$F=\dfrac{S_1^2}{S_2^2}$ 或 $F=\dfrac{S_2^2}{S_1^2}$	$F=\dfrac{較大樣本變方}{較小樣本變方}$
拒絕H_0的臨界值	$F>F_\alpha$	$F>F_{\alpha/2}$

- *例***4.8** 利用與例4.4相同之資料，零假設會被否定嗎？
  - 本例中，僅考慮兩者變異數是否在統計上相等，因此雙尾檢定較合適
  - 零假設與對立假設

$$H_o: S_1^2 / S_2^2 = 1(即\sigma_1^2=\sigma_2^2)$$
$$H_a: S_1^2 / S_2^2 \neq 1(即\sigma_1^2\neq\sigma_2^2)$$

兩母數變異數比的假設檢定

- 檢定統計量$F=2.25/0.49=4.59$
- 查表D.4得$F_{0.025,30,24}=2.21$
- 因為檢定統計量4.59=F> $F_{0.025,30,24}=2.21$，所以可拒絕H_0

□ *例4.9* 甲乙兩人持續爭論利用一台特別儀器，誰觀測角度比較精密，他們的上司聽夠後，要他們作一個測試，也就是：兩人都去觀測一特定方向，照準讀數*51*次，並各自計算變異數，測完*51*次後，乙之變異數為*0.81*，甲之變異數為*1.21*，在*0.01*顯著水準下，乙是否為較佳之觀測員？

- 本例為單尾假設檢定
- H_o： $F=S_{甲}^2/ S_{乙}^2 =1$ $(S_{甲}^2=S_{乙}^2)$
- H_a： $F=S_{甲}^2/ S_{乙}^2 >1$ $(S_{甲}^2>S_{乙}^2)$
- 檢定統計量$F=1.21/0.81=1.49$
- 查表D.4得$F_{\alpha,50,50}=1.95$
- 因為$1.49=F< F_{\alpha,50,50}=1.95$，所以不拒絕$H_o$

兩母數變異數比的假設檢定

- *例4.10* 利用*EDM*儀器，五天內重複量測一段基線，每天觀測*10*次，並取平均，求變異數後，如下表所列，在*0.05*顯著水準下，第*2*日的結果是否與第*5*日之結果顯著不同？
  - 本例為雙尾檢定
  - H_o： $S_2^2 / S_5^2 = 1$
  - H_a： $S_2^2 / S_5^2 \neq 1$
  - 檢定統計量$F=61/54=1.13$
  - 查表D.4得$F_{0.025,9,9}=4.03$
  - 因為$1.13=F< F_{0.025,9,9}=4.03$，所以不拒絕$H_o$

日	*1*	*2*	*3*	*4*	*5*
變方，S^2 (mm^2)	*50.0*	*61.0*	*51.0*	*53.0*	*54.0*

作 業

- 4.1(a)、4.2(b)、4.3(c)、4.11