Hypothesis Testing
~ Basic concept ~

假設檢定基本觀念

基本觀念

關於假設檢定

●推論統計可分為三大領域：估計、檢定、分類與選擇。

●「假設檢定」包含了兩個動作：

〝假設〞～針對想要證實的資訊做出假設；

〝檢定〞～利用樣本訊息去檢定所設立的假設是否成立。

●假設檢定是一種彼此互斥「二分法」的統計方法。

●〝假設〞事前分為「虛無假設」（null hypothesis ，H_0）與「對立假設」（alternative hypothesis ，H_1）兩種。

●〝檢定〞事後可採用常應用的「信賴區間法」、「臨界值法」、「標準檢定法」、「P值法」等方式證明。

統計推論案例

假設某同學修商用統計學的課，宣稱他可以輕易的及格過關。請問，你相信這位同學所說的話嗎？

假設來做個不記名的調查，進行隨機抽點n位同學的意見做統計：（n次抽樣試驗）

根據所得資訊，若是計算出～

●估計同意其宣稱的比例(Proportion) – **點估計**。

●估計同意其宣稱比例的可能範圍 – **信賴區間**。

●評估其所宣稱的是「隨便說說的」或「很有把握」 – **假設檢定**。

「假設」的立場

- 假設基礎：二分法（兩種互斥的決策）
- 邏輯應用：反証法
- **虛無假說**(null hypothesis ~ $\mathbf{H_0}$)
  - 欲推翻的決策
  - 主張錯誤或是希望被否決的假設
- **對立假說**(alternative hypothesis ~ $\mathbf{H_1}$)
  - 欲證實的決策
  - 主張對的假設

如何界定虛無與對立？

所以我們將檢定所要〝推翻〞的論點，放在「虛無假設」上，相對這個假設的互斥論點，就是「對立假設」，也就是我們設想所要的期待結果。

虛無假設 與 對立假設

虛無假設（null hypothesis ，H_0）

檢定者主張錯誤或希望被否定的假設。

以數學邏輯思考，虛無假設是不存在的假設，或暫時性的假設，所以稱為「虛無」。

對立假設（alternative hypothesis ，H_1）

檢定者主張對的假設。

假設檢定的過程類似數學的反證法，也就是先假設H_0是正確的，然後透過機率看是否可以否定H_0，來驗證H_1的主張是被支持的。

所以假設檢定整個計算過程，都是在檢定「虛無假設」是否可以支持否定條件，計算過程與「對立假設」無關。

假設檢定的步驟建議

第1步 根據所要研究或判斷的問題，設立兩個假設。

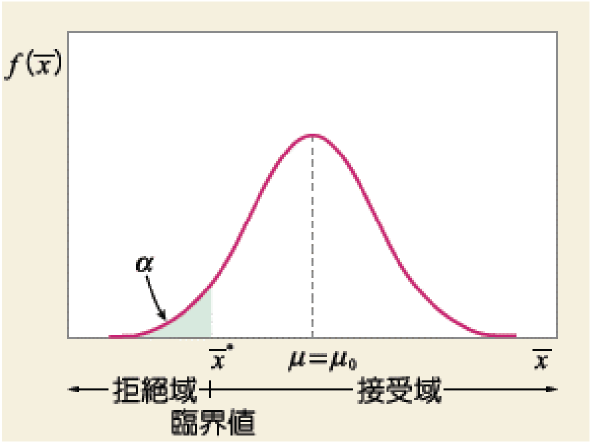
第2步 蒐集所調查的樣本資料計算「檢定統計量」。

第3步 設立顯著水準大小，建立拒絕域與不拒絕域。
（一般社會科學採用0.05作為顯著水準的標準；
醫學統計則多採用0.01作為顯著水準的標準）

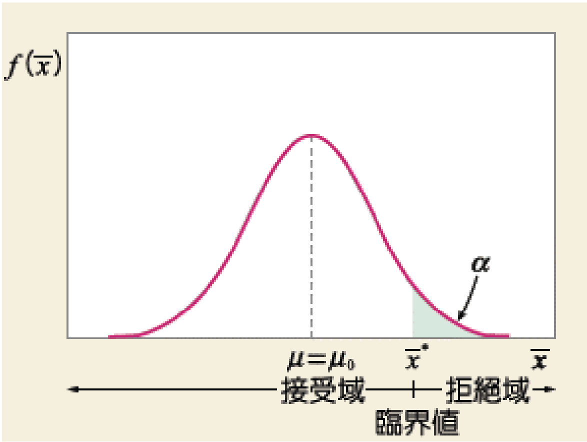
第4步 比較檢定統計量與臨界值大小，判定檢定統計量落於拒絕域或是不拒絕域內。

第5步 下結論並做決策或推論。

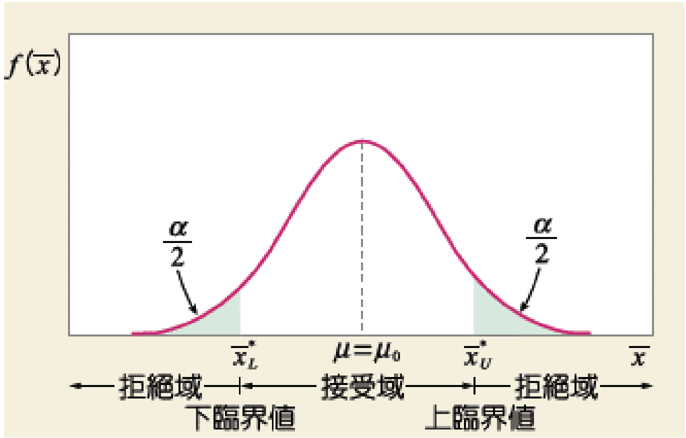
單尾檢定與雙尾檢定



左尾檢定



右尾檢定



雙尾檢定

$$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

$$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

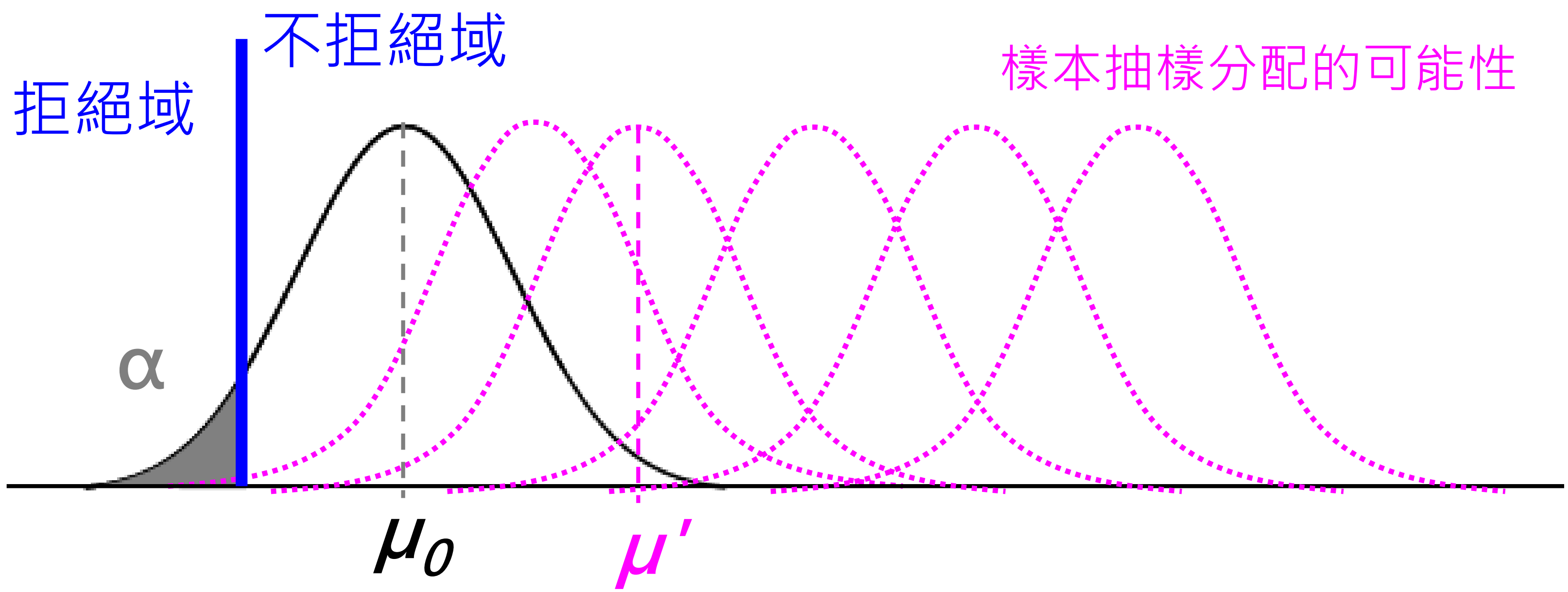
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

或 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

或 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

假設檢定整個計算過程，都是在檢定「虛無假設」是否可以支持否定條件

拒絕域與不拒絕域(接受域)



$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

或 $$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

「假設」的寫法，〝=〞一定放在「虛無假設」中。以左尾檢定為例，就是把所有大於或等於μ_0的可能情況都放在一起。若是最左邊 $\mu = \mu_0$ 被拒絕，那其他分配也等於都被拒絕，所以虛無假設也可以進行 $\mu = \mu_0$ 即可。

圖片參考：「統計學 二版」p14-5，李德治、林孟濡、童惠玲 著，博碩文化

檢定的方法（類型）

信賴區間法

在給定的信賴水準下，利用樣本統計量求出信賴區間，然後檢查此區間是否包含虛無假設值，則結論即可為拒絕虛無假設。

臨界值法

在給定的信賴水準下，先計算出臨界值，用以決定拒絕域或是不拒絕域；再檢查樣本統計量落在那個區域？進而決定拒絕或不拒絕虛無假設的方法。

標準檢定法

又稱「公式法」，是將臨界值進一步推導後所得到的方法，先求出檢定統計量，再檢查檢定統計量位於拒絕域或是不拒絕的方法。

P值法

將檢定統計量改用機率表示後，再和顯著水準比較大小，進而決定拒絕或不拒絕虛無假設的方法。

資料來源：「統計學 二版」p14-6，李德治、林孟濡、童惠玲 著，博碩文化

案例1

某飲料廠商宣稱，他們的 1 公升容器中平均至少裝有 0.98 公升的飲料。現抽出若干 1 公升裝的飲料，以檢定此一宣稱。請問，如何列「假設」議題？

案例2

某種款式的汽車平均每公升汽油可行駛 8 公里，現有一製造研究小組發明一種新的化油器系統來增加每加侖汽油可行駛的里程數。想要找到足夠的證據顯示新的化油器的確可以使里程數超過 13公里，請問，如何列「假設」議題？

案例3

在某批零件中，品管人員抽出部分為樣本，並藉以決定是否接受整批零件或因零件未符合規格而退回給供應商。假設零件規格重量平均為 750 公克，如果平均重量大於或小於 750 公克，這些零件將可能產生品質問題，請問，如何列「假設」議題？

檢定錯誤　與
檢定力函數

假設檢定的風險～檢定錯誤

Why？

當「假設」設定好後，就要利用樣本訊息來推論以進行「檢定」，作為決策拒絕或不拒絕的決定。若是抽樣過程產生偏差，就會導致決策時的誤判。

如何避免？

因為我們是透過抽樣方式來推論，所以在檢定計算時，也會瞭解可能產生誤判的機率是多少，將更多的資訊顯示出來提供給決策者有更多的判斷依據。

從統計的觀點，如果能增加抽樣的隨機性與抽取更多的樣本數，理論上，也能降低發生檢定誤判的機率。所以有時是設定誤判機率的限制條件，回頭先求取必要的樣本數，再根據可執行的抽樣方式進行抽樣調查。

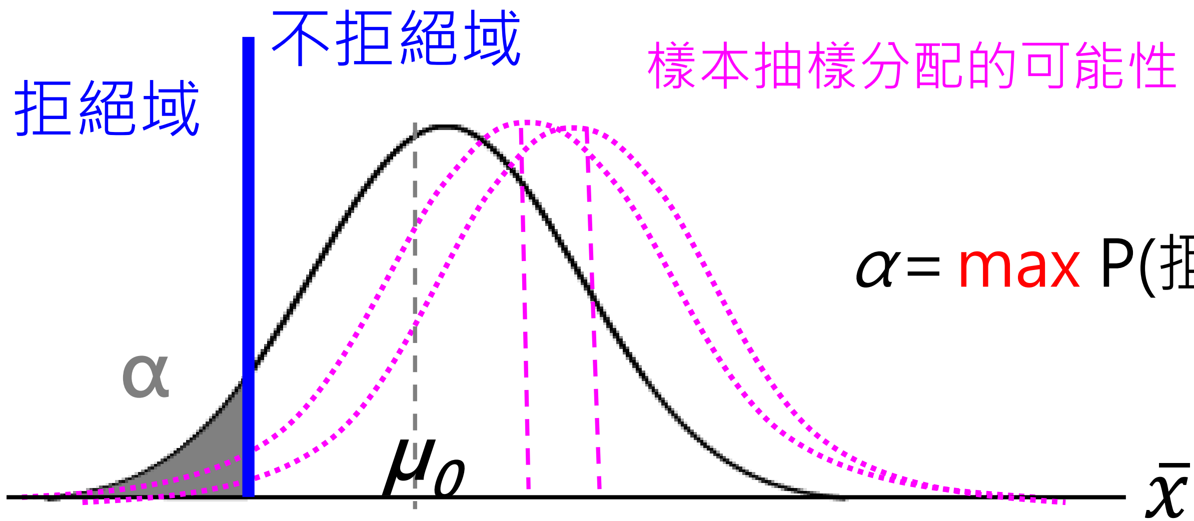
檢定錯誤的類型

真實狀況	決策	
	接受H_0	拒絕H_0
H_0為真 (H_0 True)	正確決策 機率：$1-\alpha$	型 I 錯誤 (Type I Error) 機率：α
H_0為偽 (H_0 False)	型 II 錯誤 (Type II Error) 機率：β	正確決策 機率：$1-\beta$

$\alpha = \max P$(拒絕H_0 | H_0為真)

$\beta = P$(拒絕H_1 | H_1為真)

型 I 錯誤



$\alpha =$ max P(拒絕H_0 | H_0為真)

以左尾檢定為例，「虛無假設」包含所有大於等於μ_0的一切母體可能（$\mu \geq \mu_0$）。當$\mu = \mu_0$時發生型 I 錯誤的機率最大值。

常用的 α 值是 0.05 及 0.01。

如果犯型 I 錯誤，必須付出極高的成本，則研究人員偏好較小的 α 值。只控制型 I 錯誤機率的假設檢定通常稱為顯著性檢定 (significance tests)。

圖片參考：「應用統計學 二版」p282李德治、童惠玲，博碩文化

型Ⅱ錯誤

雖然大部分的假設檢定應用都會控制型 I 錯誤的機率，但是型 II 錯誤的機率則不一定在控制中。因此，假若我們決定不拒絕 H_0，我們還是不能確定此決定有多大的信心。

統計學家通常建議我們用「不拒絕 H_0」(do not reject H_0)，而不用「接受 H_0」(accept H_0) 的陳述。

案例4

有一強烈颱風正迅速接近台灣，但不確定是否會登陸，市長需要決定明天是否放颱風假，於是他設立兩個假設，其假設為：

H_0：颱風會經過

H_1：颱風不會經過

若型Ⅰ錯誤（Type Ⅰ Error）以 α 表示；型Ⅱ錯誤（ Type Ⅱ Error ）以 β 表示，試回答下列各小題：

(1)「該放假而不放假」犯何種型態錯誤？

(2)「不該放而放假」犯何種型態錯誤？

(3)「寧可放錯假」，請問 α 增加或減少？ β 增加或減少？

資料來源：「統計學 二版」p14-32，李德治、林孟濡、童惠玲 著，博碩文化

案例4說明

<table>
<tr><th rowspan="2">真實
狀況</th><th colspan="2">決策</th></tr>
<tr><th>接受H_0(放假)</th><th>拒絕H_0(不放假)</th></tr>
<tr><td>H_0為真
(颱風會經過
→該放)</td><td>正確決策
機率：$1-\alpha$</td><td>型Ⅰ錯誤
(Type Ⅰ Error)
機率：α</td></tr>
<tr><td>H_0為偽
(颱風不會經過
→不該放)</td><td>型Ⅱ錯誤
(Type Ⅱ Error)
機率：β</td><td>正確決策
機率：$1-\beta$</td></tr>
</table>

(1)「該放假而不放假」犯何種型態錯誤？　→ 型Ⅰ錯誤

(2)「不該放而放假」犯何種型態錯誤？　→ 型Ⅱ錯誤

(3)「寧可放錯假」，請問 α 增加或減少？ β 增加或減少？

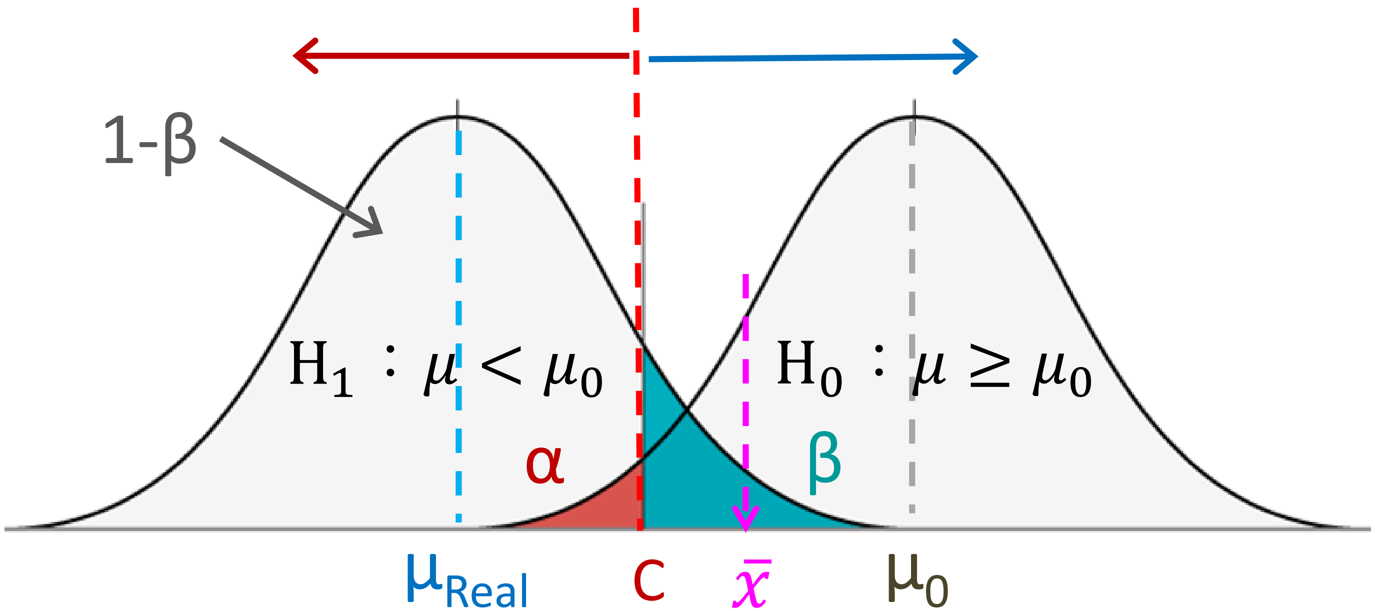
→「放錯假」表示有〝放假〞，「寧可」表示〝不該放〞，所以是「型Ⅱ錯誤」，故β會增加，而與α無關。

檢定力函數

Power of a Test

檢定力函數（power function of test）

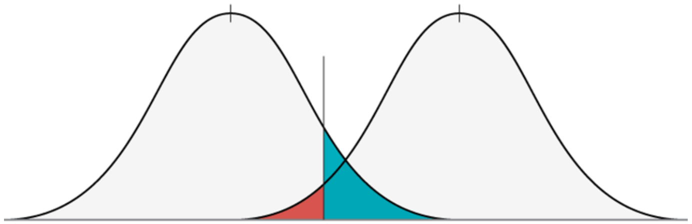
檢定力 ＝ 1-β = (接受H_1 | H_1為真)



假設實際母體的平均數為μ_{Real}（＜μ_0），若在檢定時「接受H_0」，就犯了型Ⅱ錯誤；若是要降低型Ⅱ錯誤，就是讓(1-β)所代表區域越大（更往左移或分配更集中），則β就會下降，所以在可給定（或限定）α（發生型Ⅰ錯誤機率）情況下，若β越小，就表示整個假設檢定所會犯的所有可能錯誤越小，因此把（1-β）視為檢定抽樣分配的強度，稱為「檢定力函數」。

影響檢定力因素

檢定力 = $1-\beta$ = (接受H_1 | H_1為真)



1.樣本中位數」大小：樣本數越大，檢定力越大
2.顯著水準α：一般而言α越大，檢定力越大
3.檢定統計量的選擇：一般是以平均數（$\bar{x}$）為統計量代表，若是以「中位數」為代表，因為其抽樣之標準差較大，其檢定力會較小。
4.決策法則之決定：檢定時採用左尾、右尾或雙尾亦會影響檢定力的大小。（相同條件下，單尾比雙尾更具有檢定力）

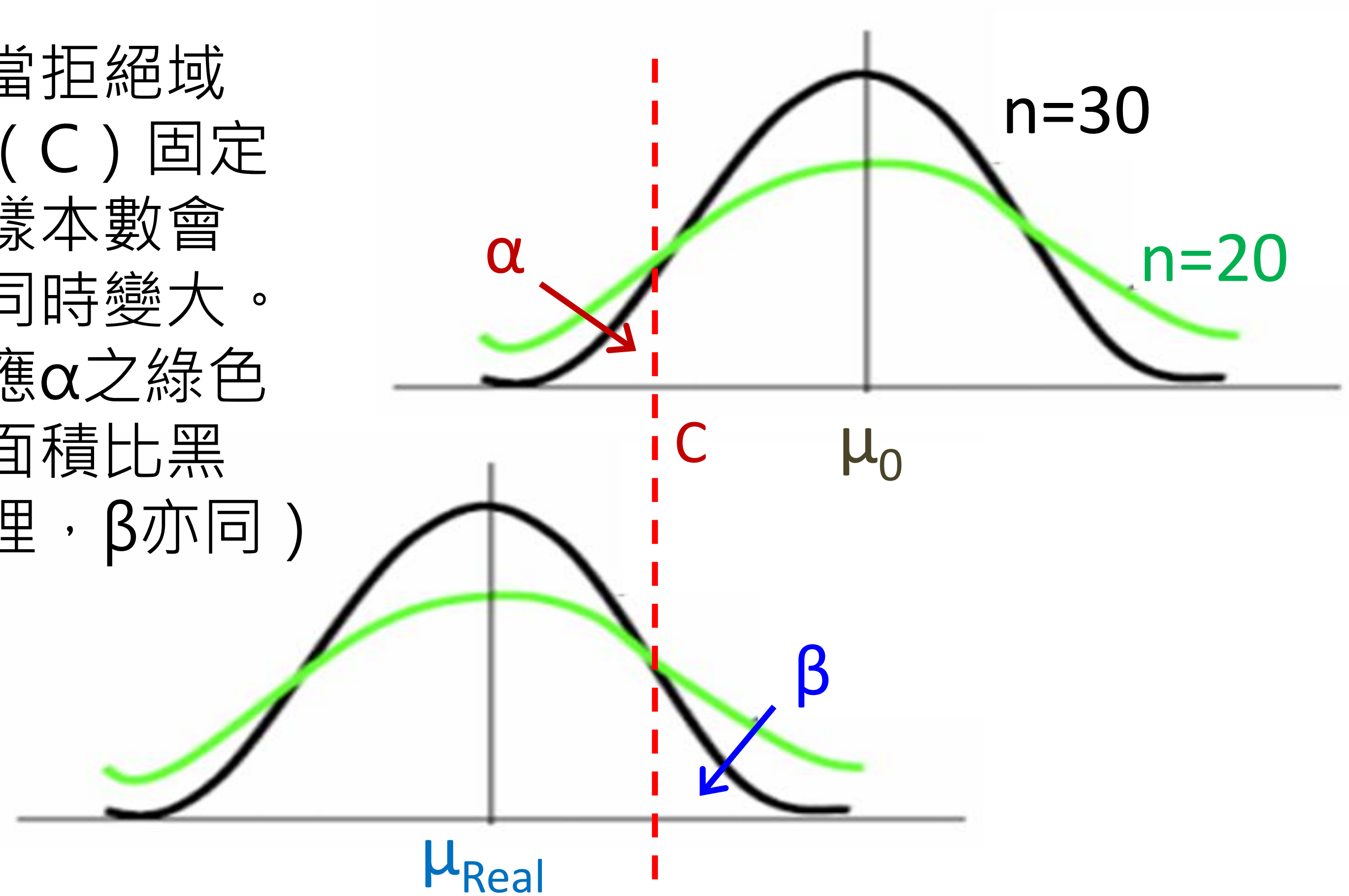
一般而言在進行檢定時，α越大β越小，有沒有可能α變大β也隨之變大？

資料來源：「應用統計學 二版」p288，李德治、童惠玲 著，博碩文化

影響檢定力因素

一般而言在進行檢定時，α越大β越小，有沒有可能α變大β也隨之變大？

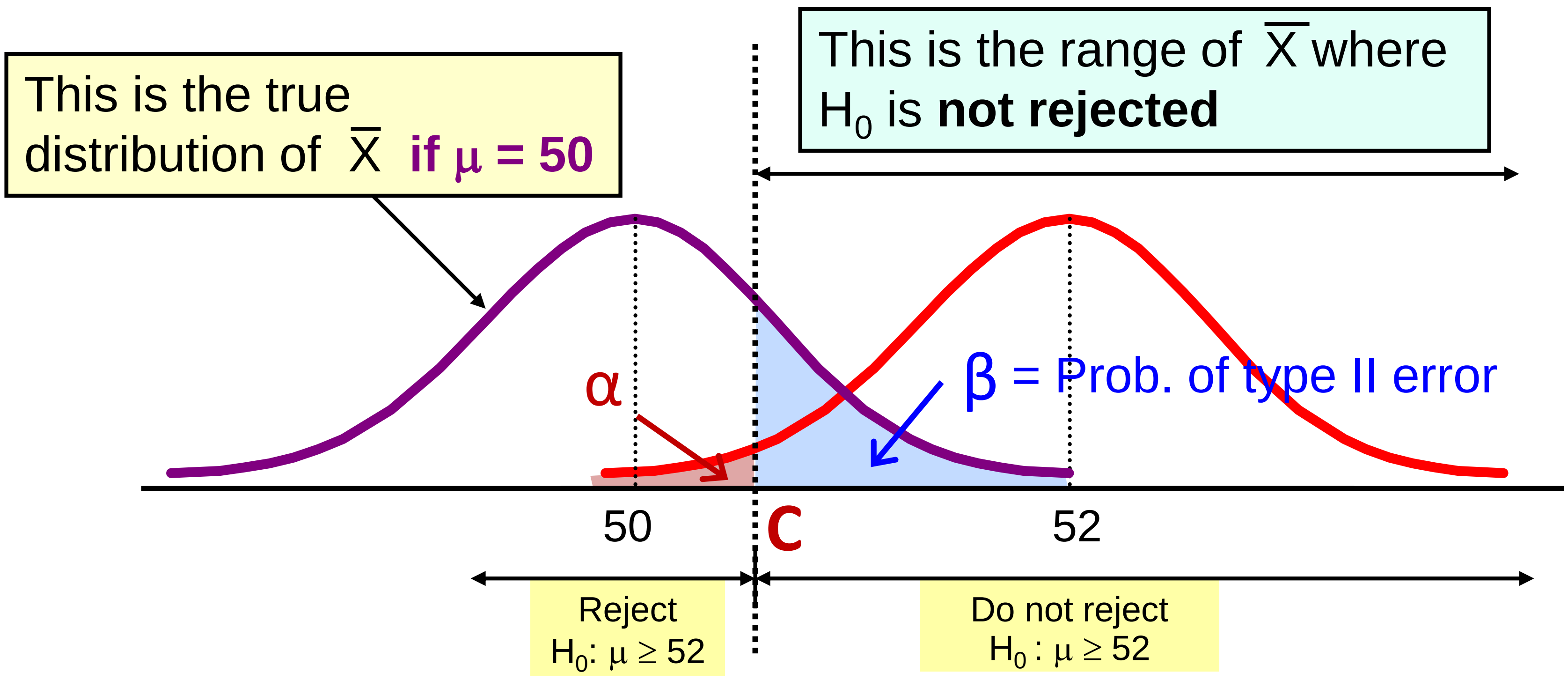
有可能，當拒絕域的臨界值（C）固定時，降低樣本數會導致兩者同時變大。（右方對應α之綠色曲線下的面積比黑色大；同理，β亦同）



如何計算檢定力？

DCOVA

- Suppose we do not reject H_0: $\mu \geq 52$ when in fact the true mean is $\mu = 50$



Here, $\beta = P(\bar{x} \geq \text{cutoff} \mid \text{if } \mu = 50)$

如何計算檢定力？

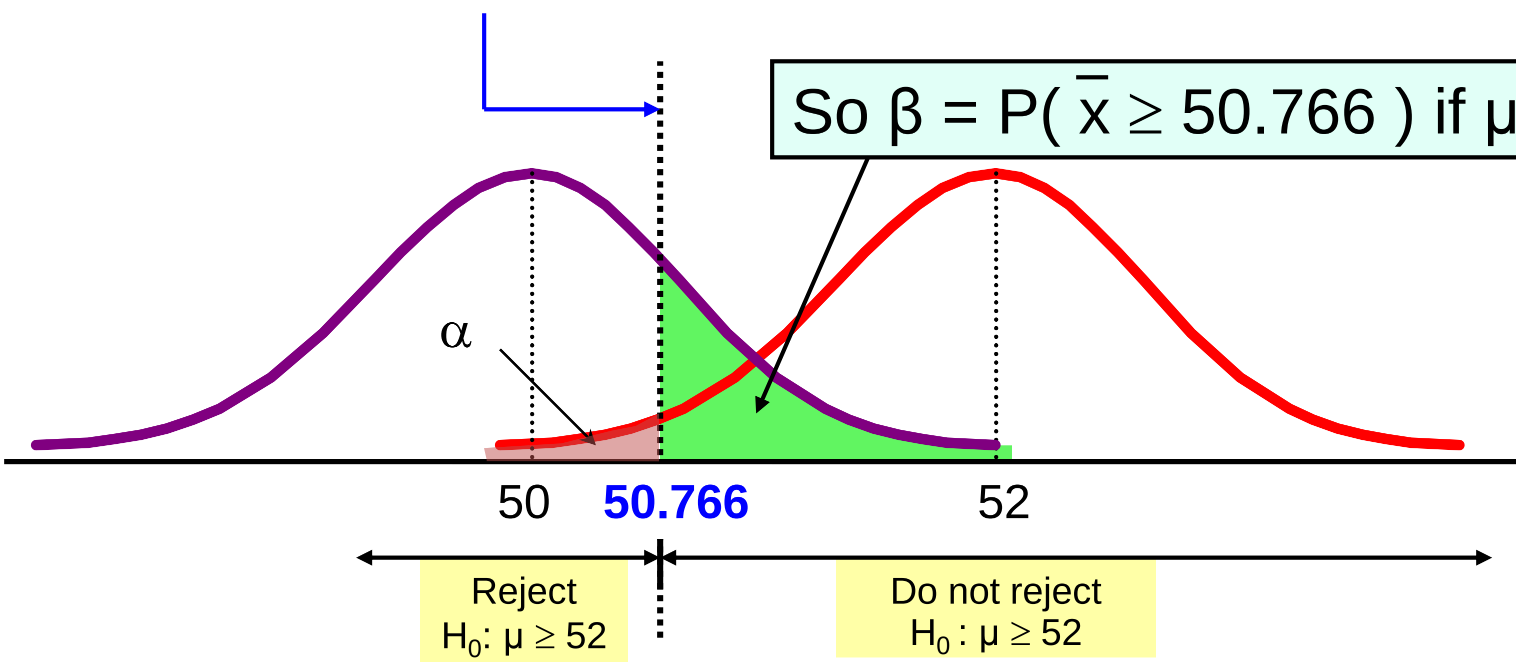
DCOVA

- Suppose n = 64 , σ = 6 , and α = .05

$$\text{cutoff} = \bar{X}_\alpha = \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} = 52 - 1.645\frac{6}{\sqrt{64}} = \boxed{50.766}$$

(for $H_0 : \mu \geq 52$)

So $\beta = P(\bar{x} \geq 50.766)$ if $\mu = 50$



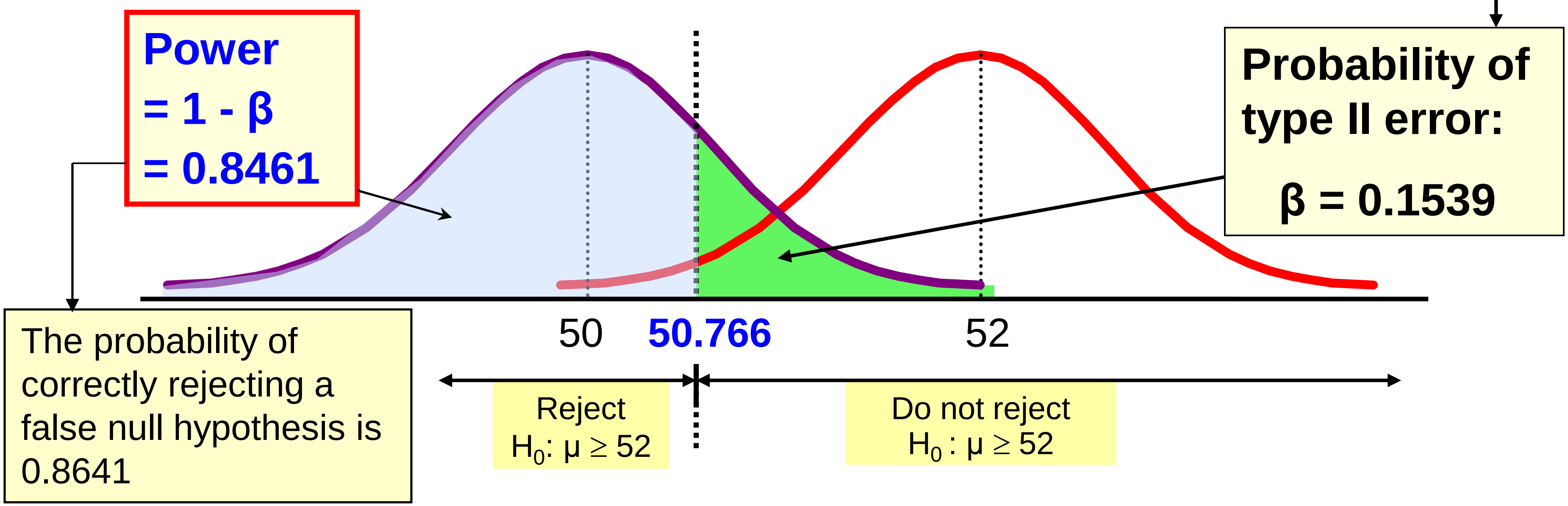
Step1：先利用α求出臨界值C

如何計算檢定力？

DCOVA

- Suppose n = 64 , σ = 6 , and α = 0.05

$$P(\bar{X} \ge 50.766 \mid \mu = 50) = P\left(Z \ge \frac{50.766 - 50}{6/\sqrt{64}} \right) = P(Z \ge 1.02) = 1.0 - 0.8461 = \boxed{0.1539}$$



Step2：計算β值

Step3：計算檢定力：power=1-β

案例5

假設$\bar{x}_4$， $\bar{x}_{25}$ 表示由常態母體N（μ，4^2）分別隨機抽取4，25個樣本之樣本平均數，欲進行μ值之檢定，若令統計假設為H_0：μ=0；H_1：μ =1，求下列二種檢定法的α與β值。

檢定法則1：若 $\bar{x}_{25} > 1.32$，則拒絕H_0：μ=0

檢定法則2：若 $\bar{x}_4 > 3.29$，則拒絕H_0：μ=0

「現代統計學」p175，吳柏林 著，五南圖書

案例5說明

(1)檢定法則1：

α = P(拒絕H_0 | H_0為真)= P($\bar{x}_{25}$ > 1.32 | H_0=0)

$= P(\frac{\bar{x}_{25}-0}{\sqrt{4^2/25}} > \frac{1.32-0}{\sqrt{4^2/25}}) = P(Z > 1.65) \fallingdotseq 0.05$

β = P(接受H_0 | H_0為偽)= P($\bar{x}_{25}$ < 1.32 | H_1=1)

$= P(\frac{\bar{x}_{25}-1}{\sqrt{4^2/25}} < \frac{1.32-1}{\sqrt{4^2/25}}) = P(Z < 0.4) = 0.66$ → 檢定力 1-β=0.34

(2)檢定法則2：

α = P(拒絕H_0 | H_0為真)= P($\bar{x}_4$ > 3.29 | H_0=0)

$= P(\frac{\bar{x}_4-0}{\sqrt{4^2/4}} > \frac{3.29-0}{\sqrt{4^2/4}}) = P(Z > 1.645) = 0.05$

β = P(接受H_0 | H_0為偽)= P($\bar{x}_{25}$ < 1.32 | H_1=1)

$= P(\frac{\bar{x}_4-1}{\sqrt{4^2/4}} < \frac{3.29-1}{\sqrt{4^2/4}}) = P(Z < 1.15) = 0.87$ → 檢定力1-β=0.13

檢定法則1的「檢定力」比法則2好

單尾檢定與雙尾檢定的檢定要點

常見用法、決策法則、檢定要點

假設檢定的類型

類別	檢定H_0項目	左尾	右尾	雙尾
單母體	平均數	$\mu \geqq \mu_0$	$\mu \leqq \mu_0$	$\mu = \mu_0$
	比例	$p \geqq p_0$	$p \leqq p_0$	$p = p_0$
	變異數	$\sigma^2 \geqq \sigma_0^2$	$\sigma^2 \leqq \sigma_0^2$	$\sigma^2 = \sigma_0^2$
兩母體	平均數差 (獨立樣本)	$\mu_1 \geqq \mu_2$	$\mu_1 \leqq \mu_2$	$\mu_1 = \mu_2$
		$\mu_1 - \mu_2 \geqq 0$	$\mu_1 - \mu_2 \leqq 0$	$\mu_1 - \mu_2 = 0$
	平均數差 (成對樣本)	$\mu_d \geqq 0$	$\mu_d \leqq 0$	$\mu_d = 0$
	比例差	$p_1 \geqq p_2$	$p_1 \leqq p_2$	$p_1 = p_2$
		$p_1 - p_2 \geqq 0$	$p_1 - p_2 \leqq 0$	$p_1 - p_2 = 0$
	變異數差異	$\sigma_1^2 \geqq \sigma_2^2$	$\sigma_1^2 \leqq \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
		$\sigma_1^2 / \sigma_2^2 \geqq 1$	$\sigma_1^2 / \sigma_2^2 \leqq 1$	$\sigma_1^2 / \sigma_2^2 = 1$

左尾檢定的檢定要點

假設檢定

$$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

給定顯著水準 α ，其對應的數值為C（臨界值）

決策法則

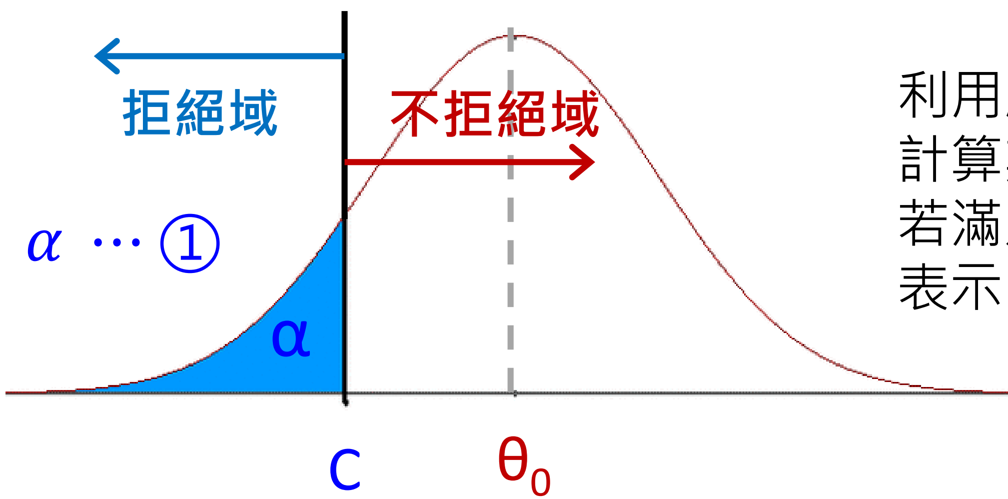
根據抽樣樣本，計算檢定統計量 θ^* ：

若 $\theta^* \geq C$ → 不拒絕 H_0

若 $\theta^* < C$ → 拒絕 H_0

抽樣分配示意圖

$$P(\hat{\theta} \leq C) = \alpha \cdots ①$$



利用所抽樣的樣本，計算其統計量 θ^* ，若滿足式①條件，即表示「拒絕H_0」

右尾檢定的檢定要點

假設檢定

$$\begin{cases} H_0: \theta \le \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

給定顯著水準 α ，其對應的數值為C（臨界值）

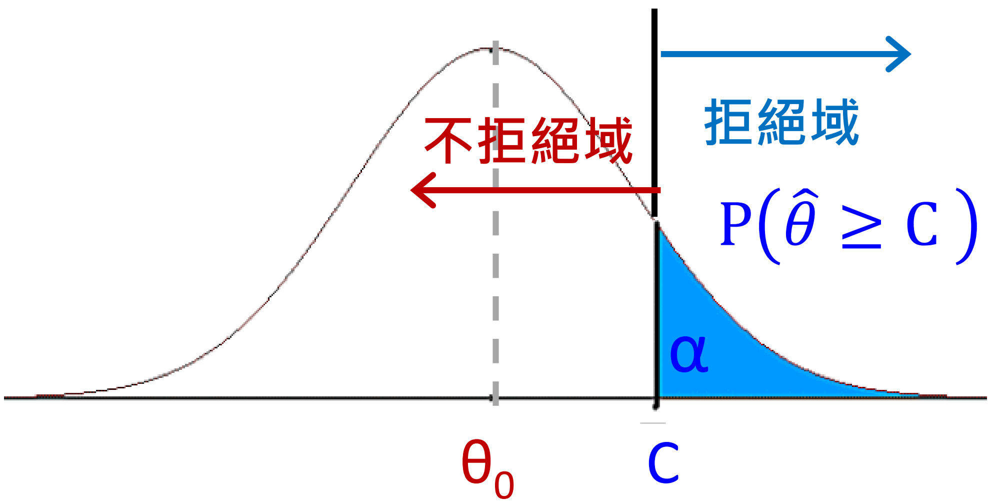
決策法則

根據抽樣樣本，計算檢定統計量 θ^* ：

若 $\theta^* \le C$ → 不拒絕 H_0

若 $\theta^* > C$ → 拒絕 H_0

抽樣分配示意圖



$$P(\hat{\theta} \ge C) = \alpha \cdots ②$$

利用所抽樣的樣本，計算其統計量 θ^* ，若滿足式②條件，即表示「拒絕H_0」

雙尾檢定的檢定要點

假設檢定

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

給定顯著水準 α，其對應的數值為C_L與C_U（臨界值的下限與上限）

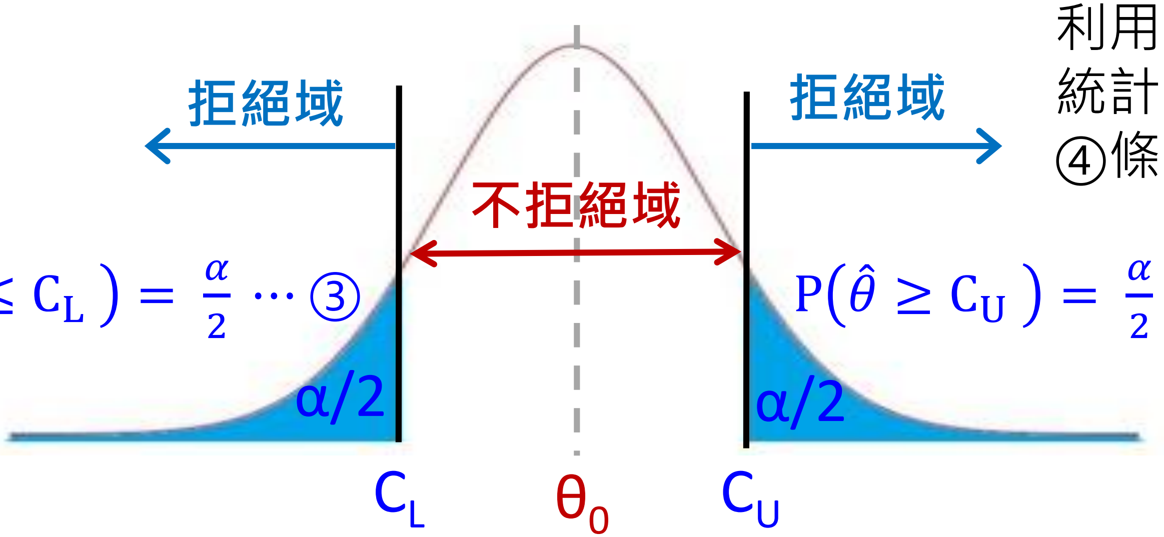
決策法則

根據抽樣樣本，計算檢定統計量 θ^*：

若$C_L \leq \theta^* \leq C_U$ → 不拒絕 H_0

若 $\theta^* < C_L$ 或 $\theta^* > C_U$ → 拒絕 H_0

抽樣分配示意圖



$P(\hat{\theta} \leq C_L) = \frac{\alpha}{2} \cdots ③$

$P(\hat{\theta} \geq C_U) = \frac{\alpha}{2} \cdots ④$

利用所抽樣的樣本，計算其統計量 θ^*，若滿足式③或④條件，即表示「拒絕H_0」

控制α、β條件下，所需的樣本數

起心動念

●抽樣時，受限抽樣方法的好壞，難免會產生判斷的錯誤。

●當型Ⅰ錯誤變小，通常會造成型Ⅱ錯誤變大。

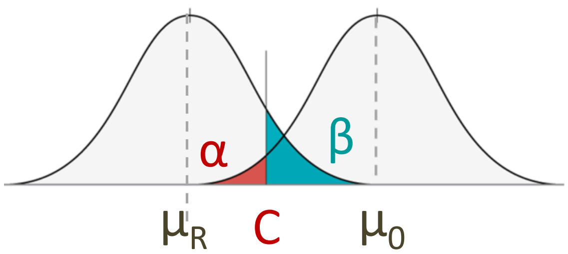
●一般型Ⅰ錯誤所造成的後果遠比型Ⅱ錯誤嚴重，因此通常會先控制型Ⅰ錯誤(α)在合理範圍後再進行檢定。

●實際上，除非知道真實母體分配資訊，也無法計算型Ⅱ錯誤(β)。若想同時減少α、β，就必須增加樣本數。

在有限的調查資源下，我們可以先試算想要控制的α與β所對應的樣本數；評估可進行的樣本數後，才進行實際的檢定統計調查。

在給定α、β值，如何計算樣本數？

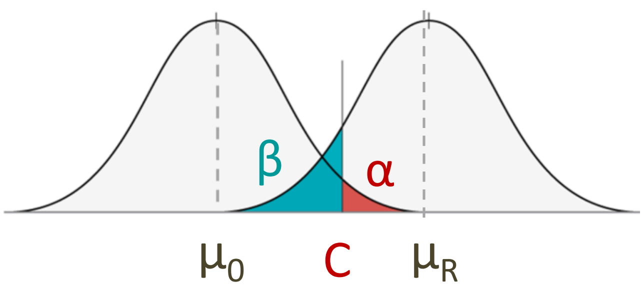
左尾檢定



$$C = \mu_0 - z_\alpha\sqrt{\frac{\sigma^2}{n}} = \mu_R + z_\beta\sqrt{\frac{\sigma^2}{n}}$$

$$\rightarrow \quad n = \frac{\sigma^2\left(z_\alpha + z_\beta\right)^2}{(\mu_0 - \mu_R)^2}$$

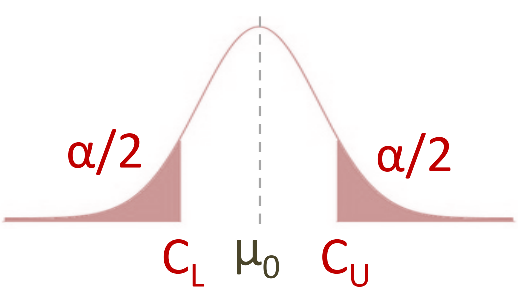
右尾檢定



$$C = \mu_0 + z_\alpha\sqrt{\frac{\sigma^2}{n}} = \mu_R - z_\beta\sqrt{\frac{\sigma^2}{n}}$$

$$\rightarrow \quad n = \frac{\sigma^2\left(z_\alpha + z_\beta\right)^2}{(\mu_R - \mu_0)^2}$$

雙尾檢定



$$C_L = \mu_0 - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} = \mu_R + z_\beta\sqrt{\frac{\sigma^2}{n}}$$

或計算C_U

$$\rightarrow \quad n \underset{C_L}{=} \frac{\sigma^2\left(z_{\alpha/2} + z_\beta\right)^2}{(\mu_0 - \mu_R)^2} \underset{C_U}{=} \frac{\sigma^2\left(z_{\alpha/2} + z_\beta\right)^2}{(\mu_R - \mu_0)^2}$$

案例6

考慮下列兩個假設：

$$\begin{cases} H_0：\mu \ \geq \ 10 \\ H_1：\mu \ < \ 10 \end{cases}$$

已知母體變異數為25，現隨機抽取120個樣本，已知母體真實的平均數是9，在顯著水準α＝0.05時，發生型Ⅱ錯誤 β=0.2912。若現在欲降低型Ⅱ錯誤至0.1，請問需再抽取幾個樣本？

「應用統計學 二版」p292，李德治、童惠玲，博碩文化

案例6說明

若給定 α = 0.05， β=0.1 下，需抽樣：

$$n = \frac{\sigma^2\left(z_\alpha + z_\beta\right)^2}{(\mu_0 - \mu_R)^2} = \frac{5^2(z_{0.05} + z_{0.1})^2}{(10-9)^2}$$

$$= \frac{5^2(1.645+1.28)^2}{(10-9)^2} = 214$$

所以還要再抽取 214 – 120 = 94 個樣本

The End

案例1：

$H_o : \mu \geq 0.98$

$H_1 : \mu < 0.98$

案例2：

$H0 : \mu \leq 13$

$H1 : \mu > 13$

案例3：

$H0 : \mu = 750$

$H1 : \mu \neq 750$