

# 什麼是檢定?

## What is a test?

### 1 檢定的意義

- 統計推論分為兩大類，一類是估計 (estimation)，一類是檢定 (test)，兩者均是由樣本的觀察對母體的參數 (parameter)，進行某種推論 (inference)
- 什麼是檢定呢? 簡單說，檢定就是回答一個 yes/no 的問題 (例如: 某甲有罪嗎?)
- 考慮以下例子: 如果檢察官或法官懷疑某甲涉貪污罪，因為他的銀行存款總額高達 7 億 (檢定統計量 test statistic)，那麼此檢定問題可以統計語言敘述之:
  - ▷  $H_0$ : 某甲清白 → 虛無假設 (null hypothesis)
  - ▷  $H_1$ : 某甲貪污 → 對立假設 (alternative hypothesis)研究者的角色相當於檢察官或法官，根據所搜集的資料來判定是否拒絕  $H_0$  (拒絕  $H_0$  表推翻無罪的假設，即判定某甲有罪)
- 此時，研究者待驗證的 (認為是對的) 陳述應放在  $H_1$ ，而其反敘述 (認為是錯的) 則放在  $H_0$ ，他期待能找到證據 (即檢定統計量夠大) 來拒絕  $H_0$ ，以支持  $H_1$
- 檢定統計量 (test statistic) 是一個被選定來協助判斷  $H_0$ ,  $H_1$  是否為對或錯的量，在上述的例子中，適當的檢定統計量可為

$Y$  = 銀行存款的總和

則判定規則可能為: 「若  $Y > 5\text{億}$   $\Rightarrow$  某甲貪污 (拒絕  $H_0$ )」

或者，也有可能出現一個較複雜的型式，例如

$$Y = \frac{\text{銀行存款的總和} - \text{被認定過大的數字}}{\text{用物價或房價水準加以正規化}}$$

則判定規則可能為: 「若  $Y > 10$   $\Rightarrow$  某甲貪污 (拒絕  $H_0$ )」

思考: 正規化代表怎樣的意義?

- 判定結果可能有正確或錯誤的情況，錯誤的情況可再分為
  - ▷ 如果該無罪卻判有罪 (不該拒絕  $H_0$  卻拒絕了)  $\rightarrow$  Type I error (型 I 錯誤)  
(例如: 明明沒貪污，但很會投資，銀行存款就是那麼多，卻被判有罪)
  - ▷ 如果該有罪卻判無罪 (應該拒絕  $H_0$  但沒拒絕)  $\rightarrow$  Type II error (型 II 錯誤)  
(例如: 明明有貪污，但很會花錢，銀行存款並不特別多，卻被判無罪)

思考: 那一種錯誤較嚴重? (I  $\rightarrow$  冤獄, II  $\rightarrow$  逍遙法外)

- 令  $\alpha = P(\text{Type I error})$ ,  $\beta = P(\text{Type II error})$

若 A 法官認定 5 億以上算貪污，B 法官認定 10 億以上算貪污

那麼，兩法官的  $\alpha$  與  $\beta$  大小關係為何？

A 法官容易將人定罪（不想讓有罪的人逍遙法外，但易造成冤獄） $\rightarrow \alpha$  大， $\beta$  小

B 法官不易將人定罪（不想造成冤獄，但易讓有罪的人逍遙法外） $\rightarrow \alpha$  小， $\beta$  大

不可能兩種機率同時變小，表示法官在「冤獄」和「逍遙法外」間必需有所取捨

Table 3.3 Classifying hypothesis testing errors and correct conclusions			
Result of test		Reality	
		$H_0$ is true	$H_0$ is false
	Significant (reject $H_0$ )	Type I error = $\alpha$	✓
	Insignificant (do not reject $H_0$ )	✓	Type II error = $\beta$

上表中:  $H_0$ : 某甲清白， $H_1$ : 某甲貪污

Q1: 那一格可以解讀為「還給某甲清白」？（該無罪判無罪）

Q2: 那一格可以解讀為「公理正義獲得伸張」？（該有罪判有罪）

Q3: 那一格可以解讀為「冤獄」？（該無罪判有罪）

Q4: 那一格可以解讀為「逍遙法外」？（該有罪判無罪）

- 檢定的原則: 先選定願意承擔的  $\alpha$  (顯著水準)，依此求出判斷基準 (臨界值  $c$ )

對上述某甲貪污例子中的法官而言: ( $\alpha = P(\text{Type I error}) \rightarrow$  造成冤獄的機率)

法官需先選定其願意承擔造成冤獄的機率  $\alpha$ ，依此決定判斷貪污的基準

A 法官選擇  $\alpha = 0.05$ ，因此決定了 5 億的認定標準

$\rightarrow$  願意承擔 5% 的誤判冤獄的機率

B 法官選擇  $\alpha = 0.01$ ，因此決定了 10 億的認定標準

$\rightarrow$  只願意承擔 1% 的誤判冤獄的機率

思考:

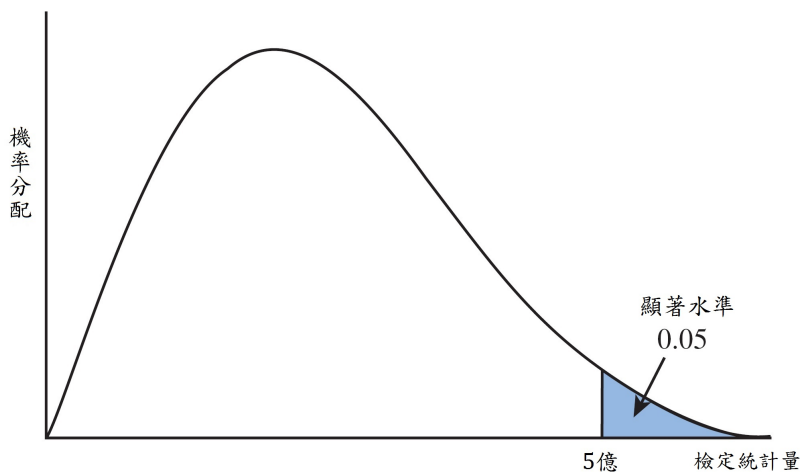
C 法官選擇  $\alpha = 0.000001$ ，決定了 1000 億的認定標準，他的想法為何？

這樣會造成什麼結果？

- 機率分配 (probability distribution) 的導入

我們常聽到  $Z$  檢定,  $t$  檢定,  $\chi^2$  檢定,  $F$  檢定等, 這些代表檢定統計量所遵循的「機率分配」, 此機率分配有何意義? 我們理解的方式是這樣的:

- ▷ 對於一個無罪的一般人來說 ( $H_0$  成立), 其檢定統計量  $Y$  遵循一個機率分配 (寫成  $Y_{\text{random}}$ )
- ▷ 對於某個特定人 (例如某甲) 來說, 其檢定統計量  $Y$  則是一個數值 (寫成  $Y_{\text{observed}}$ )



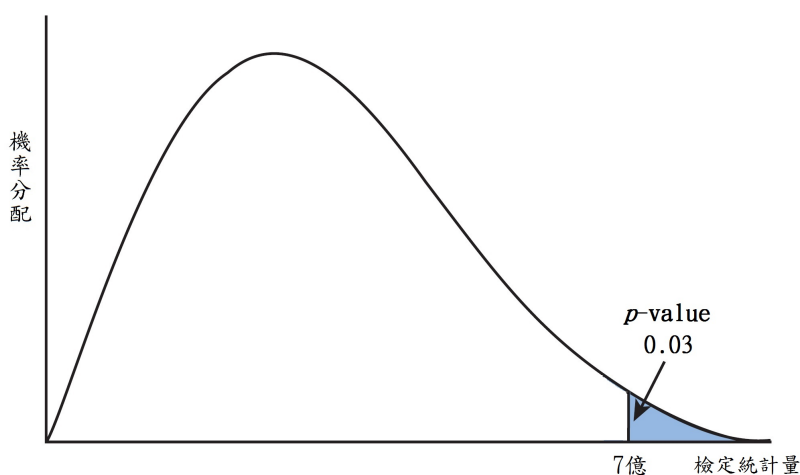
註 1: 一定要知道檢定統計量的機率分配, 才能執行檢定

註 2: 至於一個檢定統計量為何是  $Z, t, \chi^2, F$  分配, 這是數學家或統計學家的事 (從實用的角度我們可不予理會)

註 3: 法官由顯著水準  $\alpha$  所決定的臨界值 (剛才以  $c$  表示), 可寫成  $Y_{\text{critical}}$

- 統計的初學者很常在這幾個概念中搞混, 試問自己是否仍完全清楚其間差異?

$Y_{\text{random}}$  (分配) [一般人],  $Y_{\text{critical}}$  (數字) [法官: 5 億],  $Y_{\text{observed}}$  (數字) [某甲: 7 億]



- $p$ -value 的意義

$p$ -value 的正規定義是

$$p\text{-value} = P(Y_{\text{random}} > Y_{\text{observed}} \mid H_0)$$

在某甲的例子中，代表：對於一個未犯貪污罪的人來說（當  $H_0$  成立時），會看到像某甲這麼大的銀行存款金額的機率有多少？

重要觀念： $p$ -value 愈小，代表愈可能推翻  $H_0$ （無罪），愈傾向支持  $H_1$ （有罪）

實用規則： $p\text{-value} < \alpha$ （顯著水準） $\Rightarrow$  拒絕  $H_0$ （超級重要，無論如何都要記住）

## 2 兩種檢定的方法論

- 請先說服自己以下兩個敘述是正確的（理解顯著水準  $\alpha$  與  $p$ -value 的差異）

$$\alpha = P(Y_{\text{random}} > Y_{\text{critical}} \mid H_0) = P(\text{Type I error})$$

$$p\text{-value} = P(Y_{\text{random}} > Y_{\text{observed}} \mid H_0)$$

- 方法 1: 傳統法（臨界值法）

(a) 由  $\alpha$  決定臨界值  $Y_{\text{critical}}$

(b) 若  $Y_{\text{observed}} > Y_{\text{critical}}$ （或其變形）則拒絕  $H_0$

- 方法 2:  $p$ -value 法

(a) 由  $Y_{\text{observed}}$  算出  $p$ -value

(b) 若  $p\text{-value} < \alpha$  則拒絕  $H_0$ （此法不需先決定顯著水準  $\alpha$ ）

- 討論：(1) 上述兩個方法為何等價？(2) 那個方法比較好？（從考試和實用的角度）

- 討論：單尾和雙尾的差異

Q: 單尾（one-tailed）和雙尾（two-tailed）意義為何？

Q: 對方法 1（傳統法）的影響為何？

Q: 對方法 2（ $p$ -value 法）的影響為何？

- 查表練習

雖然實際上各位不會需要自己查表（又不是大學生要考試），但學會查表將有助於對檢定觀念的充份理解

▷ 使用方法 1 需要的能力：給予  $\alpha \Rightarrow$  求出  $Y_{\text{critical}}$

▷ 使用方法 2 需要的能力：給予  $Y_{\text{observed}} \Rightarrow$  求出  $p$ -value

註：查表求  $p$ -value 只能得到近似值，一般均由電腦軟體算出

### 3 結果的正確陳述

- 如果拒絕  $H_0$ ，顯示數據支持  $H_1$   
正確講法：在 5% 的顯著水準下拒絕虛無假設 (reject  $H_0$  at 5% level of significance)
- 如果無法拒絕  $H_0$ ，並不表示接受  $H_0$  (不可以說接受  $H_0$ )  
正確講法：無法拒絕虛無假設 (fail to/do not reject  $H_0$  at 5% level of significance)  
錯誤講法：接受虛無假設 (accept  $H_0$  at 5% level of significance → wrong!)
- 討論：「無法拒絕」(fail to reject) 和「接受」(accept) 有什麼不同？

### 4 常見的檢定有那些？

- 檢定常用的四大分配： $Z$ ， $t$ ， $\chi^2$  (卡方)， $F$
- 最初階的比例檢定 (基於  $Z$  分配)

	單一母體 one population	兩個母體 two populations
proportion	$H_0: p = 0.5$ $Z$ -統計量	$H_0: p_1 = p_2$ $Z$ -統計量

- 基本核心四大檢定 (用到  $t$ ， $\chi^2$ ， $F$  分配)

	單一母體 one population	兩個母體 two populations
mean	$H_0: \mu = 5$ $t$ -統計量	$H_0: \mu_1 = \mu_2$ $t$ -統計量
variance	$H_0: \sigma^2 = 10$ $\chi^2$ -統計量	$H_0: \sigma_1^2 = \sigma_2^2$ $F$ -統計量

- 思考：該如何學一個檢定？(掌握原則其實很簡單)
  - 問清楚檢定的虛無假設  $H_0$  為何？
  - 問清楚檢定統計量的公式為何？
  - 問清楚檢定統計量所遵循的機率分配為何？  
(有時會有使用上的限制，例如樣本要夠大，稍加注意即可)
  - 學會用軟體算出  $p$ -value 的方法 (不要用傳統法，也不需要查表)
  - 作一兩道例題抓到觀念 (感覺)
  - 設法連結到自己工作場域相關 (例如：風險值的穿透率檢定)