Ch5-1. 假設檢定 hypothesis testing

最常見的推論統計方法，先統計假設 (hypothesis)，再搜集資料檢驗假設 (testing)
實際問題: 欲了解台北人的智力是否高於一般人
統計作法: (1) 假設: 假設台北人智力沒有高於一般人
(2) 檢定: 隨機取樣一些台北人，進行智力測驗後，比較是否符合假設
若不符合，則不支持假設，反面論證台北人智力高於一般人
P.S. 反證邏輯 (falsification) 欲論證 A 事件為真，透過否定 A 的反面以論證
P.S. 智力測驗 $WAIS\text{-}IV \sim N(100, 15^2)$

【假設】
根據研究議題形成兩統計假設

- 虛無假設 (null hypothesis, H_0): 當操弄無效時母數的值
  上例: 「台北」此一效果無效時，台北人 WAIS-IV 將不高於一般人
  寫法: $H_0: \mu \leq 100$ (需包含等號)
- 對立假設 (alternative hypothesis, H_1): 虛無假設的反面 (通常為研究者欲求)
  上例: 「台北」效果有效時，台北人 WAIS-IV 高於一般人
  寫法: $H_1: \mu > 100$ (也可以寫 H_a)

【檢定】or 考驗
搜集資料以檢驗資料是否支持虛無假設

檢定概念說明:
若抽樣 9 人，(情況 1) 9 人平均 $\bar{X} = 110$；(情況 2) $\bar{X} = 100.1$，兩者皆大於 100 (H_0)
但我們更願意相信前者真有超過 100，後者可能僅因抽樣時的誤差造成
檢定便是決定樣本統計量與虛無假設的差距是否足夠大，而超過誤差能解釋的範圍

- 抽樣分配 (sampling distribution)
  定義: 若對母群進行無數次樣本數固定 (n) 的取樣 (取後放回)，每次取樣時皆計算樣本統計量 (如: 樣本平均數)，得無數統計量所形成的分配稱為該樣本統計量的抽樣分配 (如: 樣本平均數的抽樣分配)
  功能: 抽樣分配用來決定樣本統計量的「抽樣誤差」
- 中央極限定理 (central limit theory, CLT)
  定義: 給定平均數為 μ 變異數為 σ^2 的母群，其樣本平均數的抽樣分配之平均數為 $\mu_{\bar{X}} = \mu$，變異數又稱為變異誤為 $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ (標準誤則為 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$)，且當樣本數 n 越大時樣本平均數的抽樣分配越趨近常態分配 (不論母群分配為何)
  功能: 用以描述樣本平均數的抽樣分配長什麼樣子，可將「抽樣誤差」具體量化

【假設檢定邏輯】
(1) 假設虛無假設成立
(2) 於虛無假設成立下，建立樣本統計量的抽樣分配的數學模型
(3) 根據中央極限定理推估此抽樣分配特性 (平均數，變異誤，機率分配)
(4) 將手邊樣本與此抽樣分配進行比較
若樣本不符合此抽樣分配 (樣本發生機率過低)，則拒絕虛無假設，否則保留
P.S. 顯著水準 (significant level, α): 發生機率低於此預設水準時，拒絕虛無假設

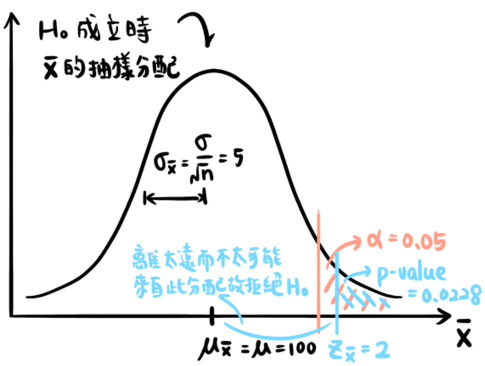
【舉例】
欲了解台北人的智力是否高於一般人，故取樣 9 位台北人進行智力測驗得以下
$X = \{110, 127, 117, 99, 110, 82, 100, 114, 131\}$ (例 ch5_1)

$\boxed{\text{Step1.}}$ 確立 H_0 & H_1
$\begin{cases} H_0: \mu \leq 100 \\ H_1: \mu > 100 \end{cases}$，其中 μ 代表台北人母群智力平均

$\boxed{\text{Step2.}}$ 決定顯著水準 (通常定在 0.05)
$\alpha = 0.05$

$\boxed{\text{Step3.}}$ 計算樣本統計量 (比較平均時，應計算樣本平均數)
$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{110+127+\cdots+131}{9} = 110$

$\boxed{\text{Step4.}}$ 於虛無假設成立下，建立樣本平均數的抽樣分配的模型



根據 CLT，此樣本平均數的抽樣分配之
平均數為 $\mu_{\bar{X}} = \mu = 100$，(此處留意 H_0 的 μ 為多少)
標準誤則為 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$

$\boxed{\text{Step5.}}$ 比較手邊實際樣本與抽樣分配差異 (以 z 分數計算樣本具抽樣分配中心點多遠)
$z_{\bar{X}} = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{110-100}{15/\sqrt{9}} = 2$ (樣本平均距 H_0 中心點往右 2 個標準誤)

$\boxed{\text{Step6.}}$ 比較樣本發生機率與顯著水準 (查詢 z 表)
$p.value = P(z \geq z_{\bar{X}}) = P(z \geq 2) = 0.0228 < 0.05 = \alpha$ (距 2 個標準誤的機率僅 0.0228)

$\boxed{\text{Step7.}}$ 寫下統計決定及質性解釋
$\because p.value < \alpha \quad \therefore reject\ H_0$
證據顯示台北人的智力高於一般人

P.S. 灰底是寫在考卷上的部分

【重要概念】

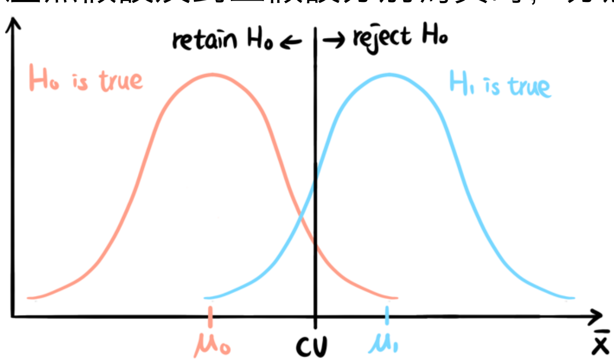
- α & β

型 I 誤差 (type I error, α): H_0 為真時，錯誤拒絕 H_0 的機率，$P(reject\ H_0 \mid H_0 = T)$
型 II 誤差 (type II error, β): H_1 為真時，錯誤保留 H_0 的機率，$P(retain\ H_0 \mid H_1 = T)$
統計檢定力 (power, $1-\beta$): H_1 為真時，正確拒絕 H_0 的機率，$P(reject\ H_0 \mid H_1 = T)$

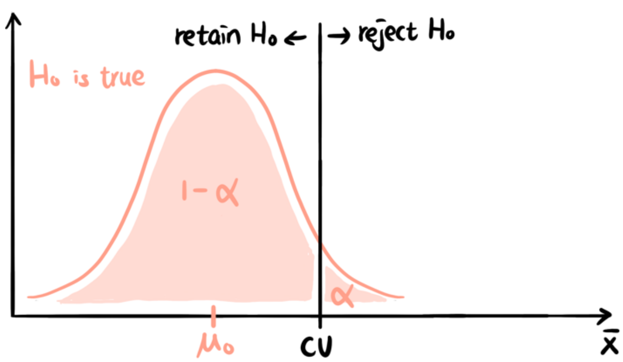
		真實情形	
		H_0 為真	H_1 為真
統計決策	$reject\ H_0$	α (type I error)	$1-\beta$ (power)
	$retain\ H_0$	$1-\alpha$	β (type II error)

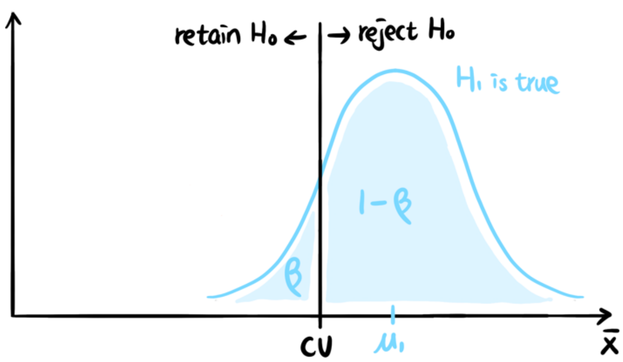
舉例: type I error: 台北人智力實際上未高於一般人，卻錯誤決策其高於一般人的機率
type II error: 台北人智力實際上高於一般人，卻錯誤決策其未高於一般人的機率

虛無假設及對立假設分別為真時，分別建構兩個樣本平均數抽樣分配如下圖



分別看兩抽樣分配下面積與臨界值的關係，可區分出上述四種條件機率，如下兩圖





- 臨界值 (critical value, CV)
  定義: 用以與樣本統計量或檢定值比較，以決策是否拒絕虛無假設的臨界值

  決斷: 當樣本統計量超過 CV，則 $reject\ H_0$，否則 $retain\ H_0$

  此例: 當 α 設定為 0.05 時 (面積為 0.05)，查詢 z 表可得對應 z 值為 1.645
  此時 z 臨界值 $z_{CV} = 1.645$
  表示當樣本 $z_{\bar{X}}$ (此例為 2) 超過 1.645 時則拒絕虛無假設
  也可透過 z 臨界值回推樣本平均數之臨界值 $\frac{\bar{X}_{CV}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = z_{CV} = 1.645$
  $\frac{\bar{X}_{CV}-100}{5} = 1.645 \Rightarrow \bar{X}_{CV} = 108.225$,
  表示 9 人樣本平均數 (此例為 110) 超過 108.225 即拒絕虛無假設
  P.S. $\bar{X} > \mathrm{CV}$ 之範圍 (此例 $\bar{X} > 108.225$) 又稱為拒絕區域 (reject region)

  影響: 受顯著水準 (α) 的設定影響
  α 越大時，CV 值減小 (決策線左移)，β 隨之增加
  因此統計決策時，應考量更願意承擔何種錯誤
  舉例: 藥物療效研究

$\mathrm{H_0}$: 藥物無效	(1) 當副作用很強時	(2) 當無副作用 (保健食品)
$\mathrm{H_1}$: 藥物有效	α 應小 (保守決策)	可承擔較大 α

- p-value
  定義: H_0 為真時，出現較此次樣本更極端的所有可能 (含此次取樣) 的機率總和

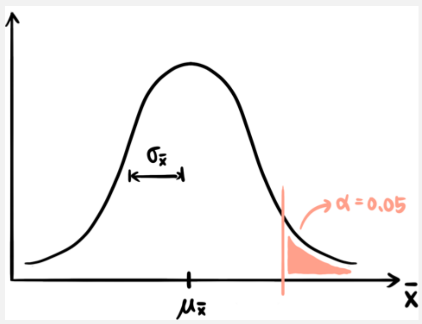
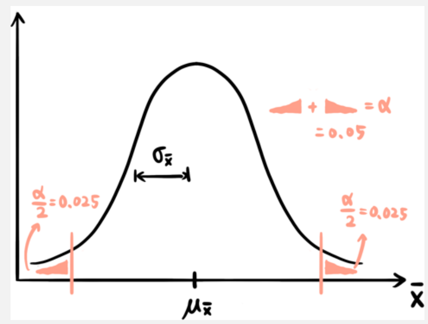
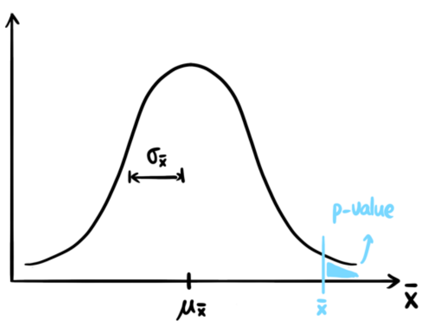
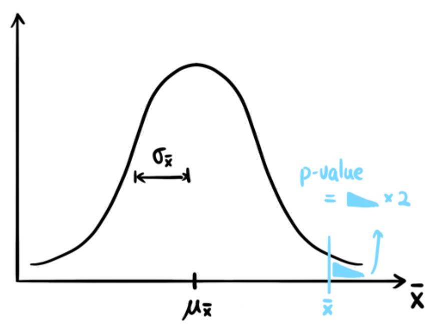
  使用: 若 $p.value < \alpha$，則 $reject\ H_0$，否則 $retain\ H_0$
  邏輯為，在 H_0 成立下，要取到這個樣本的可能性太低，而認為 H_0 不合理

  此例: (見 Step4 的圖) 9 人樣本之平均數 ($\bar{X} = 110$)
  於 H_0 為真時之抽樣分配下，距離中心 2 個標準誤 ($z_{\bar{X}} = 2$)
  較此樣本更極端的所有可能 位於 2 個標準誤之外 ($z_{\bar{X}} = 2$ 的右側範圍)
  據定義，$p.value = P(z \geq 2) = 0.0228$ (查 z 表之 z 為 2 時的右側面積)

- 針對 H_0 決策
  所有統計決策，不論接受或拒絕，皆是針對 H_0
  因為「假設檢定」中的「假設」，是假設 H_0 成立，「檢定」則是檢驗此假設
  P.S. 並無 $reject\ H_1$ 或 $retain\ H_1$ 的寫法

- 避免使用「證實」等肯定詞語
  $reject\ H_0$ 僅說明樣本與 H_0 的差異來自誤差的可能性低，但仍可能犯錯 (拒絕時的犯錯機率為 α)，因此用語應以「證據支持」等非論證假設的用語

- 檢定的方向性 (單尾 or 雙尾)

	單尾 (one-tailed)	雙尾 (two-tailed)
舉例	想了解台北人智力是否高於一般人 ⇒ 「高於」具有方向性	想了解台北人智力是否不同於一般人 ⇒ 「不同於」沒有方向性
檢定寫法	$\begin{cases} H_0: \mu \le 100 \\ H_1: \mu > 100 \end{cases}$	$\begin{cases} H_0: \mu = 100 \\ H_1: \mu \neq 100 \end{cases}$
拒絕區 (α 相同 為 0.05)		
p-value		

- 顯著 (significant) 的意義

顯著 (拒絕虛無假設) 表示樣本統計量與虛無假設差異明顯，無法單由誤差解釋！！但不表示差異很大！！

觀察 p-value 計算: $p.value = P(z \ge z_{obs}) \Rightarrow z_{obs} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

當 樣本平均數與 μ 差異越大 或 樣本數增加時 都會使 p 值降低 (顯著性增)

效果量 (effect size) 的意義

用以反應研究操弄的程度大小，或變項間關聯強度

觀察效果量計算: $Cohen's\ d = \frac{\mu_1-\mu_0}{\sigma}$

樣本數不影響效果量計算

【再舉例】二項分配假設檢定
舉例：賭場宣稱自家硬幣公正 (正反面出現機率相同為 0.5)，測試一枚硬幣投 50 次有 30 次正面，請問此硬幣是否公正？

Step1. & Step2.

$\begin{cases} H_0: \pi = 0.5 \\ H_1: \pi \neq 0.5 \end{cases}$，其中 π 為硬幣真實的正面機率，$\alpha = 0.05$

方法一

Step3. 計算樣本統計量 (比較「機率」時，應計算樣本正面「機率」)

$p = \frac{30}{50} = 0.6$

Step4. 於虛無假設成立下，建立樣本機率的抽樣分配的模型

假設 H_0 成立時，X 為此硬幣擲 50 次出現正面次數，則 $X \sim B(n,p) = B(50,0.5)$

根據二項分配趨近常態分配

$X \sim B(n,p) \equiv N\big(np, np(1-p)\big) = N(25,12.5)$

$\Rightarrow p = \frac{X}{n} \sim N\left(p, \frac{p(1-p)}{n}\right) = N(0.5,0.005)$

Step5. 比較手邊實際樣本與抽樣分配差異 (以 z 分數計算樣本距抽樣分配中心點多遠)

$z_p = \frac{p-\pi}{\sigma_p} = \frac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.6-0.5}{\sqrt{0.005}} \doteqdot 1.4142$

Step6. 比較樣本發生機率與顯著水準 (or 比較 z 臨界值)

$p.value = 2 * P\big(z \geq z_p\big) = 2 * P(z \geq 1.4142) \doteqdot 0.1573 > 0.05 = \alpha$

or $z_p \doteqdot 1.4142 < 1.96 = z_{\alpha/2}$

Step7. 寫下統計決定及質性解釋

$\because p.value > \alpha \quad \therefore retain\ H_0$

這個硬幣還算公正

方法二

Step3. 於虛無假設成立下，直接計算樣本 p-value

假設 H_0 成立時，X 為此硬幣擲 50 次出現正面次數，則 $X \sim B(n,p)$

$p.value = 2 * P(X \geq 30) = 2 * \sum_{i=30}^{50} C_i^{50}\, 0.5^i (1-0.5)^{50-i} \doteqdot 0.2026 > 0.05 = \alpha$

Step4. 寫下統計決定及質性解釋

$\because p.value > \alpha \quad \therefore retain\ H_0$

這個硬幣還算公正

P.S. 若 n 很大時，通常選擇方法一，因為方法二的組合數會算到瘋掉

Ch5-2. 估計 estimation

搜集資料並計算一數值或數值區間，用以估計母數，是另一推論統計方法

【點估計】point estimation
定義: 計算樣本統計量 (ex: $\bar{X}$) 以估計母群母數 (ex: μ)

【區間估計】interval estimation
定義: 計算一數值區間以估計母群母數

- 信賴區間 (confidence interval, C.I.)
  定義: $1-\alpha\ C.I.$ 透過樣本與設定之 α 計算區間，此計算進行無數次得無數信賴區間，其中有 $1-\alpha$ 比例次數之信賴區間會涵蓋母數
  計算: $1-\alpha\ C.I. = \bar{X} \pm z_{\alpha/2} * \sigma_{\bar{X}}$
  (其中 $z_{\alpha/2}$ 會隨著抽樣分配改變)
  此例: 計算台北人智力之 95 信賴區間 (以 9 人樣本計算)
  $95\ C.I. = 110 \pm z_{0.025} * 5 = 110 \pm 1.96 * 5 = [100.2, 119.8]$
- 95 信賴區間與雙尾假設檢定有等價關係 (see 證明 5)

【再舉例】二項分配之信賴區間
舉例: 2023 年 TPP 內参民調 **1082** 人，其中支持柯侯配者佔 **44%**，求 **95** 信賴區間

解題: (此處需理解區間估計及二項分配趨近常態分配此兩概念)
根據區間估計公式 $95\ C.I. = p \pm z_{0.05/2} * \sigma_p$
此處 p 之估計為 0.44，尚須知曉 p 之抽樣分配以估計 σ_p

假設 X 為 1082 人中支持柯侯配之人數，X 即符合二項分配
$X \sim B(1082, 0.44) \equiv N(1082, 1082 * 0.44 * 0.56)$
$\Rightarrow p = \frac{X}{1082} \sim N\left(0.44, \frac{0.44*0.56}{1082}\right)$
$\Rightarrow \sigma_p = \sqrt{\frac{0.44*0.56}{1082}} = 0.0151$

因此 95 信賴區間為 $p \pm z_{\frac{0.05}{2}} * \sigma_p = 0.44 \pm 1.96 * 0.0151 = 0.44$ **± 0.0296**
$= (41.04\%, 46.96\%)$

媒體識讀: (上述資料整理為媒體常用語如下)
此次民調時間於…，調查方式為…、對象為…、抽樣方式為…，其中共 **1082** 份有效樣本，於 **95%** 信心水準下，有 **44%** 民眾支持柯侯配，且抽樣誤差為**正負 2.96 個百分點**