

# 掌握 AB test 的精隨 - 假設檢定的進階概念與種類



簡報閱讀



範例與作業



問題討論



學習心得(完成)

重要知識點

假設檢定的種類

以 H1 所定訂範圍分類

以樣本的範圍分類

檢定目的下，假設檢定的種類 (以雙樣本為例)



重要知識點



- 掌握假設檢定的種類，釐清你想了解的所

## 假設檢定的種類

假設檢定的種類，可以透過三種面向來區分。

根據  $H_1$  所定訂範圍分類，可分為

- 右尾檢定
- 左尾檢定
- 雙尾檢定

根據樣本的範圍，可分為

- 單樣本檢定
- 雙樣本檢定
- 成對樣本

根據檢定目的，可分為

- 平均數檢定
- 比例檢定

## 以 $H_1$ 所定訂範圍分類

以單樣本為例，說明右尾檢定、左尾檢定與雙尾檢定，要選擇哪一種以你所想要了解的問題或角色而定。

	左尾檢定	右尾檢定	雙尾檢定
使用情境	一位店長認為其品牌在市場之佔有率至多為12%，一共訪問了300為消費者，其中有31位表示喜歡。	一位店長認為其品牌在市場之佔有率至少為12%。	一位店長認為其品牌在市場之佔有率為12%。
假設	$H_0: \mu \geq 0.12$ $H_1: \mu < 0.12$	$H_0: \mu \leq 0.12$ $H_1: \mu > 0.12$	$H_0: \mu = 0.12$ $H_1: \mu \neq 0.12$

	單樣本檢定	雙樣本檢定 (非相依)	成對樣本檢定 (相依)
資料特性	做實驗只分一組	做實驗分成兩組	分成兩組，但兩組有前後或相依的特性。
使用情境	一位店長認為其品牌在市場之佔有率至多為12%，一共訪問了300為消費者，其中有31位表示喜歡。	想比較男生女生在薪資上是否有差異性？ 父親每日陪伴孩子時間和母親每日陪伴孩子的時間是否有差異？	成對樣本： 分析一群夫妻，夫與妻分別的年收入多寡是否有差異。  重複量測： 參加減肥試驗的一群人，參加試驗前與規律運動3個月後的體重是否有差異。
使用時機	檢樣我們收集的樣本，所算出的統計值，是否高於、低於或等於某一特定值	需比較兩種群體，或者兩種選擇下，哪一個選擇或群體較好	

## 檢定目的下，假設檢定的種類 (以雙樣本為例)

	平均數檢定	比例檢定
使用情境	台灣男性的平均腰圍是否比女性的平均腰圍來的多？	兩種不同的email主旨，50封是統一式開頭，50封是個人化開頭，請問個人化的開信率有比統一化的開信率來的高？
平均數檢定	雙樣本 t 檢定 $H_0: \mu_{女} - \mu_{男} \geq 0$ $H_1: \mu_{女} - \mu_{男} < 0$	雙樣本 t 檢定 $H_0: p_{統} - p_{客} \geq 0$ $H_1: p_{統} - p_{客} < 0$

## 樣本的範圍下，假設檢定的種類

	單樣本檢定	雙樣本檢定 (非相依)	成對樣本檢定 (相依)
使用情境	一位店長認為其品牌在市場之佔有率至多為12%，一共訪問了300為消費者，其中有31位表示喜歡。	想比較男生女生在薪資上是否有差異性？	成對樣本： 分析一群夫妻之中，夫與妻分別的年收入多寡是否有差異？
平均數檢定	單一樣本 t 檢定 $H_0: \mu \geq 0.12$ $H_1: \mu < 0.12$	獨立樣本 t 檢定 $H_0: \mu_{女} \geq \mu_{男}$ $H_1: \mu_{女} < \mu_{男}$	成對樣本 t 檢定 $D = X_{夫} - X_{妻}$ $H_0: D \leq 0$ $H_1: D > 0$

## 統計上重要的抽樣分配

- 還記得先前教的離散性分配、連續型分配，在假設檢定中，做了實驗(抽樣)後，收集了一堆樣本，描述這些樣本的分配，稱為抽樣分配。
- 統計上重要的抽樣分配有
  - t 分配
  - 卡方分配
  - F 分配
- 這次的課程中，將介紹 A/B test 常用的 t 分配和卡方分配。

### t 分配前的先備知識 - 卡方檢定

卡方分配是標準常態的變形，Z 為標準常態分配，Z 的平方為自由度為 1 的卡方分配 →

$$Z \sim N(0,1), Y = Z^2, Y \sim \chi^2(1)$$

卡方分配具有以下特性：

1. N 個卡方分配相加為自由度為  $nv$  的卡方分

$$\rightarrow Y_i \sim \chi^2(v) \rightarrow \sum_{i=1}^n Y_1 + \dots + Y_N \sim \chi^2(nv)$$

2. N 個常態分配的平方相加為自由度為  $n$  的卡方分配

$$Z_i \sim N(0,1) \rightarrow \sum_{i=1}^n Z_i^2 + \dots + Z_n^2 \sim \chi^2(n)$$

### 何謂 t 分配

稱為 t 分配

- 應用於在母體標準差( $\sigma$ )未知的情況下，不論樣本數量大小皆可應用學生 t 分配估計呈常態分布且變異數未知的總體的平均值。

## t 分配和標準常態分配的定義

假設  $X$  是常態分配的隨機變數，平均數為  $\mu$ ，標準差為  $\sigma$ ，抽出  $X_1, \dots, X_n$  服從  $X \sim N(\mu, \sigma^2)$

樣本平均為：

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

樣本變異數為：

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

標準常態

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

由於  $\sigma$  未知

用  $s_n$  估計  $\sigma$

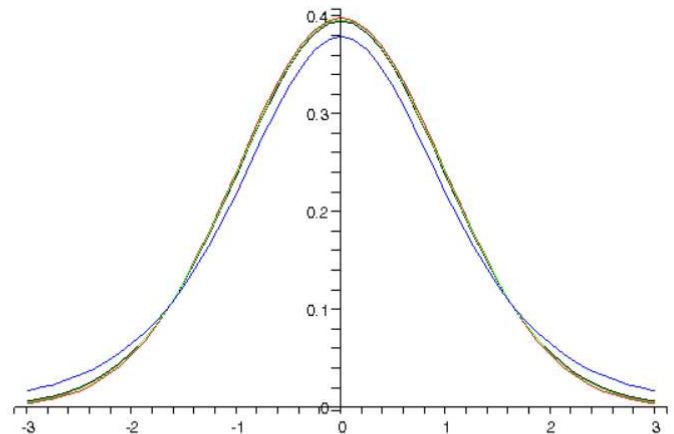
t 分配

$$T = \frac{\bar{X}_n - \mu}{\frac{s_n}{\sqrt{n}}}$$

## t 分配與常態分配的關係

t 分配當樣本數越多時，自由度就會越高，從右圖模擬觀察到，t 分配當自由度為 30 時，很接近常態分配。

藍色為  $t$  分配機率密度函數(自由度=5)  
 黑色為  $t$  分配機率密度函數(自由度=30)  
 綠色為  $t$  分配機率密度函數(自由度=50)



自由度為5的  $t$  分配 圖形較為平坦(藍)  
 自由度為30的  $t$  分配 圖形(黑)與常態分配(紅)已經接近重合。  
 自由度為50的  $t$  分配 圖形(綠)雖然更加接近常態分配(紅)  
 但與自由度為30的  $t$  分配 圖形(黑)差別 很小

圖形來源：

[http://www.cust.edu.tw/mathmet/t\\_and\\_normal.htm](http://www.cust.edu.tw/mathmet/t_and_normal.htm)

## Z 分配和 t 分配在檢定的使用時機

母體標準差	樣本數量	選用分配
已知	小樣本	Z 分配
已知	大樣本	Z 分配
未知	小樣本(小於30)	t 分配
未知	大樣本	t 分配 / Z 分配 (兩者皆可)

在母體標準差未知的情況下，不論樣本數量大小或大皆應選用學生t檢定

母體標準差通常都是未知居多，因此在假設檢定時統稱為 t 檢定，大樣本時可以採用 Z 分配方便計算。

## 假設檢定的誤差類型

時，檢定誤差也會比較高，發生檢定錯誤是難免的，但是我們希望誤差可以在我們可以容忍的範圍內，根據檢定結果與真實結果，我們可以區分成下面四個象限。

		檢定後的決定	
真實			(O) True Positive
		(O) True Negative	

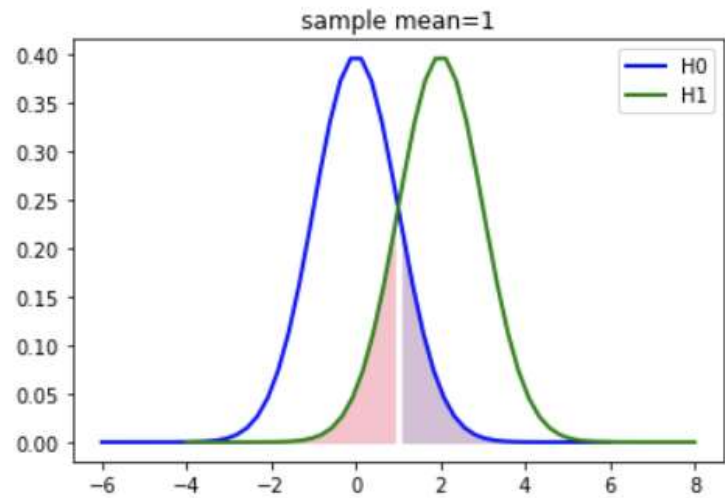
- $\alpha$  : Type I error/型一誤差，又稱**偽陽性 false positive**， $H_0$  是對的，但是我們做了實驗後，卻拒絕 $H_0$ ，又稱**顯著水準 (significant level)**，設定  $\alpha$  值愈小，表示希望檢測時的誤判機率愈低(即希望檢定能愈準確)
- $\beta$  : Type II error/型二誤差，又稱**偽陰性 false negative**， $H_0$  是錯的，但是我們做了實驗後，卻沒有證據拒絕  $H_0$ 。
- $1 - \beta$  : 又稱檢定力， $H_0$  是錯的，但是我們做了實驗後拒絕  $H_0$  的能力。

若用驗孕棒為一位未懷孕的女士驗孕，結果是已懷孕(positive)，這是第一型錯誤。

若用驗孕棒為一位已懷孕的女士驗孕，結果是未懷孕(negative)，這是第二型錯誤。

檢定  $H_0 : \mu < 0$ ， $H_1 : \mu > 0$ ，假設在  $H_0$  為真下，分配為  $N(0,1)$ ， $H_1$  為真下，分配為  $N(2,1)$ 。

$\alpha = p(\text{拒絕}H_0|H_0\text{為真})=p(\bar{X} > 1|H_0\text{為真}) \rightarrow \text{紫色}$   
 $\beta = p(\text{接受}H_0|H_1\text{為真})=p(\bar{X} < 1|H_1\text{為真}) \rightarrow \text{粉色}$



## 知識點回顧

- 掌握假設檢定的種類，分成三種面向，包含根據  $H_1$  所定訂範圍分類、樣本的範圍與根據檢定目的。
- t 分配與卡方分配是統計上重要的抽樣分配，用在檢定上。

假設檢定的誤差類型，包含兩種

- 型一誤差，又稱**偽陽性 false positive**
- 型二誤差，又稱**偽陰性 false negative**

## 延伸閱讀

型 I 錯誤、型 II 錯誤與 p 值

網站：[科學online](#)

可以閱讀看不同的角度講解 誤差類型與 p 值。



若  $H_0$  (虛無假說) 為真, 但結論卻否決  $H_0$ , 則犯了第一型錯誤, 而稱犯第一型錯誤的機率為第一型錯誤率 (Type I Error Rate), 其發生的機率以  $\alpha$  表示, 或稱顯著水準 (significant level)。

## 二、型 II 錯誤 (Type II Error) :

若  $H_1$ （對立假說）為真，但結論卻接受  $H_0$ ，則犯了第二型錯誤，而稱犯第二型錯誤的機率為第二型錯誤率 (Type II Error Rate)，其發生的機率以  $\beta$  表示。另外，統計上常稱  $1 - \beta$  為檢定力 (Power) (圖一)。

事實 (Truth)		
決策 (Decision)	$H_0$ 為真	$H_1$ 為真
無法棄卻	決策正確	型 II 錯誤
$H_0$	1 - 一個	$b$
(拒絕) $H_0$ )		(II型錯誤)
棄卻	型 I 錯誤	決策正確
$H_0$	一種	1 - $\beta$
(拒絕 $H_0$ )	(類型I錯誤)	

### 下一步：閱讀範例與完成作業