

# 掌握 A/B test 的精隨 - 假設檢定的概念



簡報閱讀



範例與作業



問題討論



學習心得(完成)

重要知識點

設計網頁的選擇的難題

無所不在的 A/B test 在醫療

無所不在的 A/B test 在行銷



重要知識點



- A/B test 為什麼重要？
- A/B test 成功案例
- A/B test 是什麼？

# 虛無假設與對立假設

# p-value

## 設計網頁的選擇的難題

- 設計網頁時，只能猜測使用者的體驗來設計介面，但會因為每個人的價值觀的不同，而沒有準確的依據。

採用原設計

A

採用另一設計

B

- 有些問題的答案是永遠都討論不出來的，倒不如進行小規模的實驗，來更快地找出答案。
- A/B test 是一個幫助大家做決定的方法，讓每一個選擇不再靠感覺。

## 無所不在的 A/B test 在醫療

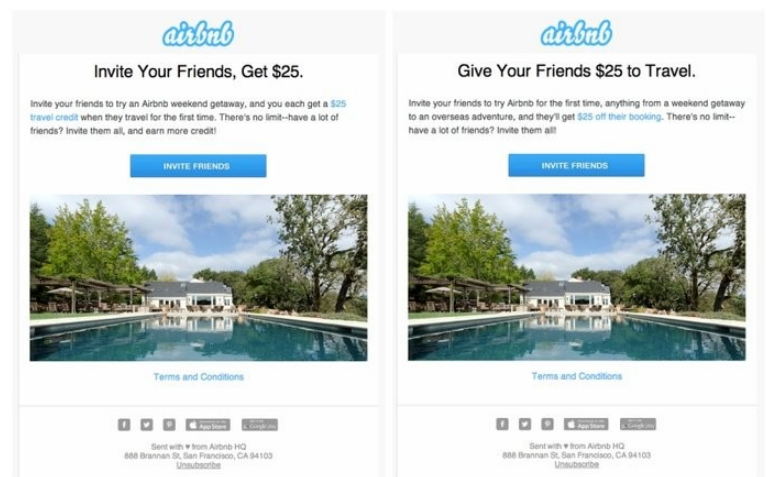
在醫療產業的 A/B test，稱做臨床試驗，隨機分派設計早已是現代醫學研究不可或缺的方法。

以下是新藥上市前的臨床試驗的流程圖，主要包含從患者的加入試驗，以及怎麼分配患者到不同的實驗組與對照組的流程設計的過程。



## 無所不在的 A/B test 在行銷

在 Airbnb 也常使用 A/B test，  
例如邀請朋友後，得到一優惠的表達方式，  
也會影響邀請朋友的訊息，  
左方是「邀請好友可以獲得25美元」，右方是  
「給你的好友贈送25美元的旅行經費」，  
你覺得哪一個你會比較有意願邀請朋友加入體驗  
Airbnb？



圖片來源：[appadhoc](#)

## 什麼是A/B test？

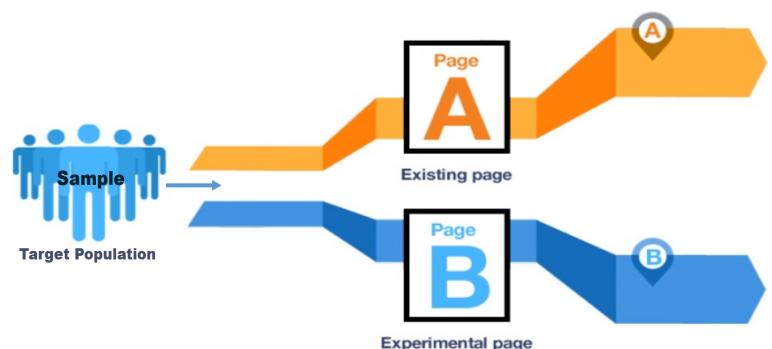
“In marketing and business intelligence, A/B testing is a term for a **randomized experiment** with **two variants, A and B**, which are the control and variation in the controlled

the technical term, **two-sample hypothesis testing**, used in the field of statistics."

- 萃取上段敘述，A/B 測試主要包含三個核心概念：「隨機化的實驗」、「一個變因，兩種選擇」、「兩種樣本的假設檢定」。
- 其實實務上應用，也不只A/B 兩種選擇，也可以A/B/C/..等。

## 什麼是隨機化的實驗？

- 隨機的重要性在於『控制住人類可控制的任何因素後，隨機化實驗能分析其因果關係』。
- 如何做：透過 seed 來產生亂數表，把每一個人任意分到某些群體中。



資料來源：<https://zhuanlan.zhihu.com/p/34301270>

## 一個變因，兩種選擇或多個選擇

- AB test 的重點要找出因果關係，一次只能允許幾個因素有變化，盡量不要太多，通常

- 一個網頁首頁的設計，這個變因涵蓋太多變動因素，包含顏色、按鈕中的文字內容與首頁排版格式都會影響，因此通常做實驗，都取一變數來看，比如只觀察使用者對於首頁的顏色偏好？顏色的選則可能有分紅色、紫色與黃色等。

## 假設檢定 - 源英國上流社會喝下午茶時的遊戲

1920 年代的劍橋大學，一群人優閒地享受下午茶時光，有位女士突然提到，先把茶加進牛奶裡，與先把牛奶加進茶裡，這兩種沖泡方式所泡出的奶茶口味截然不同，實驗中準備了 10 杯奶茶，有些是先放茶再加牛奶，有些先放牛奶再加茶，並將這些奶茶隨機排序讓這位女士品茗。

在設計實驗時，為了避免許多不相關的因素影響這位女士的口味辨別，還需要將茶和牛奶充分混合的時間、泡茶的時間及水的溫度控制一樣等等 (其他因素都控制的一樣)。

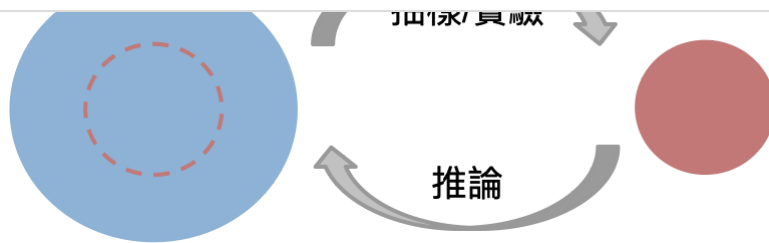
據說後來的實驗結果，這位女士真的能分辨出每一杯茶，且完全答對，結論就是下午茶的調製順序對風味有很大的影響。

資料來源：[agron.ntu.edu.tw](http://agron.ntu.edu.tw)

8 杯奶茶

4 杯牛奶加入紅茶 (樣本 1)

4 杯紅茶加入牛奶 (樣本 2)



### 母體 真相？

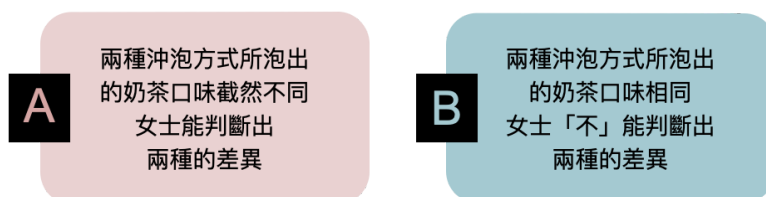
- 這兩種沖泡方式所泡出的奶茶口味是相同或不同？

### 樣本 現象

- 這位女士真的能分辨出每一杯茶，且完全答對

得到全對的結果，就能推論到真相是，兩種沖泡方式所泡出來的奶茶口味是不相同的？

## 假設檢定 Step 1：將抽象情況數值化/把抽象化的現象數量化



以  $p$  表示女士每杯說對的機率， $X$  表示 8 杯說對的杯數，則  $X$  從  $B(8, p)$  分布，透過經驗判斷， $p$  值要大於 0.7 才算女士真的有辨識能力，所以當  $p < 0.7$  就代表兩種沖泡方式所泡出的奶茶口味截然不同；

$p > 0.7$  就代表兩種沖泡方式所泡出的奶茶口味相同；

透過隨機變數，把抽象化的現象數量化，從 A/B 情境轉成  $H_0 / H_1$ 。

$p > 0.7$  就代表女士能判斷出兩種差異；

分別對應到假設檢定的  $H_0$  與  $H_1$

- $H_0$  稱為虛無假設
- $H_1$  稱為對立假設

喝茶實驗的情境下， $p < 0.7$  為  $H_0$ ， $p > 0.7$  為  $H_1$ 。

**假設檢定 Step 2：假設  $p=0.7$  值下，我們看到實驗狀況的可能性？**

- $p = 0.7$  為你設定的臨界值
- $P(X = 8) = 0.7^8 = 0.082$
- 這時你會說，有 8% 的機率那女士會剛剛好猜中，你怎麼能篤定？

**假設檢定 Step 3：增加實驗數量，泡了 20 杯，女士也全都猜對**

- $p = 0.7$  為你設定的臨界值
- $P(X = 20) = 0.7^{20} = 0.0008$
- 這時機率夠小了嗎？
- 當機率夠小了，則就能推論這兩種沖泡方式所泡出的奶茶口味截然不同。

**\* 為什麼當機率夠小了，則就能推論出結果？**

- 假設你哥哥 (P) 平日都開紅色車子(Q) [P Q]
- 你能推論出下面哪一個結果？

Q1：開紅色車子的 (Q) 就是你哥哥 (P) [Q $\rightarrow$ P]	X
Q2：不是你哥哥 (-P) 就沒有開紅色車子 (-Q) [-P $\rightarrow$ -Q]	X
Q3：沒有開紅色車子的 (-Q)，一定不是你哥哥 (-P) [-Q $\rightarrow$ -P]	O

若  $P \geq Q$  是對的（成立），我們無法保證  $Q \geq P$  或  $-P \geq -Q$  一定是對的，僅能肯定  $-Q \geq -P$  的敘述也是正確的。

## 假設檢定 Step 4：推論結果的理論基礎-反證法

Step2：

$H_0$ ：  $p < 0.7$  就代表女士不能判斷出兩種差異

$H_1$ ：  $p > 0.7$  就代表女士能判斷出兩種差異

- 回想檢定 Step 2，假設  $p = 0.7$  值下 (當  $H_0$  為真)，代表女士不能判斷出兩者的差異的假設前提下，20 杯都會猜中的機率很小 (不可能發生)。
- $P$  為女士不能判斷出兩者的差異，所以  $Q$  為不可能全部猜中。
- 但實驗的結果為全部猜中， $-Q$ ，所以  $-P$ 。
- 得到的結論為： $p > 0.7$  就代表女士能判斷出兩種差異。
- 由於  $P \rightarrow Q$ ，只能推論到  $-Q$  則  $-P$ ，代表你想證明的事情，需要放置在  $H_1$ ，而非  $H_0$ 。
- 因為我們透過邏輯判斷，推論出有足夠的證據證明  $H_0$  不是真的，來呈現  $H_1$ 「非常可能」對

## 假設檢定 Step 5：計算 p 值，是否有足夠的證據證明 $H_0$ 不是真的



上沒有差異，卻因誤差或偶發而產生資料差距的機率』。

- p值 (p-value) 在  $H_0$  為真下，比我們收集到的比我們資料還更極端的機率。
- 通常 p 值計算出來會跟顯著水準相比，顯著水準的選擇通常為 0.1、0.05、0.005 都可，就看領域上的大家的嚴謹度來選擇，p 值越小代表越嚴謹。

## 假設檢定 Step 6：p 值和顯著水準比較，決定接受或拒絕 $H_0$

- 以仕女喝茶為例，假設顯著水準設為 0.05，只做了 8 杯的實驗， $p\text{-value} = P(X \geq 8) = P(X = 8) = 0.7^8 = 0.082$ ， $0.082 > 0.05$ ，因此沒有足夠的證據，說明兩種沖泡方式所泡出的奶茶口味截然不同。
- 以仕女喝茶為例，假設顯著水準設為 0.05，只做了 20 杯的實驗， $p\text{-value} = P(X \geq 20) = P(X = 20) = 0.7^{20} = 0.0008$ ， $0.0008 < 0.05$ ，因此有足夠的證據，說明兩種沖泡方式所泡出的奶茶口味截然不同。

## 知識點回顧

A/B test 是一個幫助大家做決定的方法，讓每一個選擇不再靠感覺。

A/B 測試主要包含三個核心概念：「隨機化的實驗」、「一個變因，兩種選擇」、「兩種樣本的假設檢定」。

- Step2：計算  $H_0$  為真實，我們看到實驗狀況的可能性
- Step3：確認實驗的狀況是否罕見
- Step4：運用反證法，進行推論
- Step5：計算 p 值
- Step6：p 值和顯著水準比較，決定接受或拒絕  $H_0$

## 延伸閱讀

隨機分配好重要

產品經理實際案例分享，為什麼隨機分配很重要！



資料來源：[medium](https://medium.com/@michaelshen/why-randomization-is-important-for-product-managers-8f8f8f8f8f8f)

## 常見隨機分配的方法 (Variety of randomization)

隨機分配的方法，會根據資料特性不同，而採用不同的方式，簡單隨機分配是最常見的一種，還有很多不同的變形，可以透過延伸閱讀，更了解不同方式的適用性。

Stratified randomization：分層隨機	依照特定變因先進行分層，行樣本再於層內進行隨機分配，之後再將各組分派後的結果整併為治療組與控制組
Cluster randomization：分群隨機	執行時須先將樣本區分成數個組內個體相似性高的群組 (cluster)，接著再以這些 cluster 而非個體，來進行隨機分派。

資料來源：[statisticbyjerry.blogspot](http://statisticbyjerry.blogspot)

[下一步：閱讀範例與完成作業](#)

