

專題: Comparing Two Population Means: A Simple Simulation Practice

目的:

本專題針對具備統計學基礎與應用的學生，利用基礎統計學中簡單的「兩母體平均數差異」的檢定問題，練習統計研究中最常運用的模擬技術「Monte Carlo Method」¹，做為統計研究領域的初探。

為什麼:

統計方法的應用與研究是兩回事，一個將統計方法應用在實務資料上，另一個針對實務問題提出解決的統計方法。一般而言，在大學階段的統計學習較著重於理論的理解與方法的應用，對於學術研究的過程鮮少碰觸。每個可以被用來解決問題的統計方法除了必須在理論上被證明可行外，通常還需要透過大量模擬資料的驗證，方能說服使用者，且運用這個方法所得到的推論才有意義。本單元利用解決一個簡單且眾所周知的問題，讓初學者初嚐研究中常見的模擬過程，祈能順利進入統計研究的殿堂，成為指導老師的好幫手。

做什麼:

統計學者提出一個統計方法，固然需要理論根據，但實務上的效果更是重要。所謂真金不怕火煉，每個方法都要經過多方試煉，才能證實為可行的統計方法。但問題在於實務上我們常不能提供足夠的真實資料做為測試的依據，為在短時間內提供其效力證明，符合問題假設的各種模擬資料便取而代之。本單元要求初學者利用電腦程式模擬符合問題假設的各種資料，並且大量地重複執行，讓該統計方法的「統計現象」可以呈現出來。以假設檢定而言，我們關心一個檢定統計式對於型一誤的維持及其檢定力的表現。

¹對於Monte Carlo Method 不熟悉者可以上網查 Wikipedia 的說明。

怎麼做:

初學者必須先瞭解「兩母體平均數差異檢定」的來龍去脈，並依循統計學中解決該問題的方式，依各種可能的假設，譬如，母體是否為常態、兩母體間的關係是否獨立、樣本數大小、兩母體變異數是否已知或相等...，分別寫出檢定統計式，並模擬出各種不同的資料，每種資料都需要大量的樣本，作為重複檢定的資料，透過大量的檢定結果，觀察檢定統計式的型一誤及其檢定力的表現。模擬的過程有幾件事要做：

- 大量模擬資料的產生 (包括符合 H_0 與 H_a 各種假設的資料)，譬如模擬兩組樣本 (x_1, x_2) ，各有 n_1, n_2 筆來自 $N(0,1)$ 的常態資料。在模擬的需求下，我們必須產生這樣的樣本 (x_1, x_2) 共 N 筆，以便執行 N 次的檢定，紀錄 N 次拒絕或接受 H_0 的檢定結果。
- 計算 N 筆兩組樣本 (x_1, x_2) 的統計量 t_1, t_2, \dots, t_N 並觀察這 N 個統計量的分佈情況 (是否與理論上對檢定統計式分配的推論一致)。
- 型一誤 (α) 的維持情況: 在給定型一誤 α 下進行檢定，如果拒絕 H_0 的資料有 N_1 筆，我們想知道錯誤拒絕的比例 N_1/N 是否接近 α 值。
- Power 的優劣評估: 同上，當資料來自 H_a 時， N_1/N 代表檢定力，越接近 1 越好。

既是模擬，次數 N 當然操之在己，一般而言越大越好，不過當 N 大到某個程度，其結果將趨於穩定，過大只是徒增計算時間。所以模擬次數的多寡只求適當即可，譬如 $N = 50000$ 次。

背景:

在統計的應用裡，我們常需要知道兩組樣本的均數是否相等。以假設檢定的方式，可以表達為

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

這是在基本統計學會學到的應用問題，雖然是基本的理論，在應用上卻非常的廣泛。檢定假設的方法通常根據虛無假設 H_0 ，建立一個檢定統計式 (test statistic)，而這個統計式的主角就是樣本。因樣本數的多寡、獨立與否或其它的假設，檢定統計式可能隨之改變。在理想的情況下，我們希望該統計式服從某個已知的分配，方便判斷樣本來自 H_0 的可能性有多高。判斷的方式便是將樣本代入統計式中，得到一個統計量，根據統計式的機率分配，計算出該統計量在這個分配函數中的落點，這個落點反映出這組樣本來自該分配的可能性，這個可能性一般稱為 p 值。² p 值愈小代表樣本來自 H_0 的機會愈小，當 p 值小於預先設定的型一誤 α 時，便做出拒絕 H_0 假設的檢定結果。不過本練習的主旨並非在 p 值的計算，而是企圖利用這個簡單的檢定問題，引導學生經歷一個研究過程，奠定未來從事未知問題研究的基礎。

檢定統計式是假設檢定的一個靈魂角色，它是怎麼來的？有些檢定統計式非常單純簡潔，有的卻非常複雜繁瑣，僅是計算便令人卻步，更難想像這樣的統計式是如何怎麼生出來的？譬如與本單元相關的一個檢定統計式

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

其中 $\bar{X}_1, \bar{X}_2, s_1^2, s_2^2$ 分別代表第一與第二組的樣本平均數與變異數, n_1, n_2 是樣本數。在某些假設條件下，統計式 t 服從 T 分配，自由度 $n_1 + n_2 - 2$ 。比起許多其他統計式，上式算是相對單純的。當統計學家提出一個檢定統計式，他必須針對該統計式做模擬實驗，才能讓人相信這是個可靠的統計式。本單元希望引導學生從事類似的模擬實驗。

研究假設與限制:

²關於 p 值，此處的說明較為簡略，不熟悉者宜參考統計學課本，從定義上去琢磨它的意義。

每個研究主題都有其假設與限制,「兩母體平均數的檢定」有哪些假設與限制呢? 不同的假設或限制會造成什麼影響呢? 譬如, 在小樣本的情況下, 當母體不為常態卻仍以常態假設的統計式做檢定, 其型一誤是否能維持? 檢定力會變差嗎? 萬一找不到統計式服從那個分配, 該怎麼辦呢? 唯有透過徹底的資料模擬, 才能對所有的假設與限制有深刻的瞭解, 才能體會後續的許多統計研究到底在做些什麼, 統計學課本的描述只能點到為止。