

混合效果模型在行為科學的應用

鄭中平

心理系
成功大學

29 April 2023

Outline

- 1 基礎篇：混合模型的 What, Why & How
 - 混合效果模型是什麼
 - 為何需要混合效果模型
 - 怎麼跑混合效果迴歸模型（多層次迴歸）
- 2 不同階層的獨變項、調節與中介
 - 不同階層的獨變項
 - 調節與中介
 - 中介
- 3 延伸篇：成長模型與實驗資料
 - 個別成長模型
 - 心理學實驗資料的階層
 - 常見問題

混合效果模型

混合效果模型是什麼？

- 混合效果模型就是模型中有固定效果以及隨機效果的模型
- 混合效果模型用來處理多層次資料
- 與很多詞意義重疊：多層次模型、隨機係數迴歸、階層線性模型...
- 我們多半會用多層次模型、多層次分析這樣的詞，強調資料的特性
- 在強調有兩類效果的情境，會用混合效果模型這個詞

什麼是多層次資料？

當資料中分析單位（人）巢套在上一層單位時，就是多層次資料

- 學生在班級中（教育、教育心理學）
- 員工在公司中（管理、組織心理學）
- 病患在醫院中（醫學、臨床心理學）
- 反應在參與者中（實驗、認知心理學）

在本次工作坊中，我們將最底層稱為 level 1，依序往上數上面例子的 level 1, level 2 單位分別是什麼？

直覺的問題

- 資料可以是三層次或以上？可以。
- 就是分組資料，不能用先前方式處理嗎？
- 分析單位好像不只一種？

以前對分組資料的處理

以分析 2 個學校 (A, B) 為例，要瞭解性別對學生成績影響為例

- 分別做兩迴歸，有兩個截距，兩個斜率

針對 A 校資料， $Y = b0_A + b1_A gender + \epsilon$

針對 B 校資料， $Y = b0_B + b1_B gender + \epsilon$

- 形成一個學校的虛擬變項 D，代表 A 校 (D=0) 與 B 校 (D=1)

令 $b0_D = b0_B - b0_A$, $b1_D = b1_B - b1_A$

$Y = b0_A + b0_D D + b1_A gender + b1_D gender \times D + \epsilon$

- 當 $D = 0$, $Y = b0_A + b1_A gender + \epsilon$ ，就是 A 校
當 $D = 1$, $Y = (b0_A + b0_D) + (b1_A + b1_D) gender + \epsilon$ ，就是 B 校

2 個迴歸式，合成一個迴歸式。有 1 個截距、1 個學校效果、1 個性別效果、1 個性別與學校交互作用，總共 4 個係數要計算。

以前對分組資料的處理

以分析 3 個學校 (A, B, C) 為例，要瞭解性別對學生成績影響為例

- 分別做三迴歸，有三個截距，三個斜率
- 形成兩個學校的虛擬變項 $D1$, $D2$ ，代表 A 校 ($D1=0$, $D2=0$)、B 校 ($D1=1$, $D2=0$) 與 C 校 ($D1=0$, $D2=1$)
$$Y = b_{0A} + b_{0D1}D1 + b_{0D2}D2 + b_{1A}gender + b_{1D1}gender \times D1 + b_{1D2}gender \times D2 + \epsilon$$
- 當 $D1 = 0, D2 = 0$, $Y = b_{0A} + b_{1A}gender + \epsilon$ ，就是 A 校
當 $D1 = 1, D2 = 0$,
 $Y = (b_{0A} + b_{0D1}) + (b_{1A} + b_{1D1})gender + \epsilon$ ，就是 B 校
當 $D1 = 0, D2 = 1$,
 $Y = (b_{0A} + b_{0D2}) + (b_{1A} + b_{1D2})gender + \epsilon$ ，就是 C 校

3 個迴歸式，合成一個迴歸式。有 1 個截距、2 個學校效果、1 個性別效果、2 個性別與學校交互作用，總共 6 個係數要計算。

以前對分組資料的處理

以分析 100 個學校 (A, B, C,) 為例，要瞭解性別對學生成績影響為例

- 分別做 100 迴歸，有 100 個截距，100 個斜率
- 形成 99 個學校的虛擬變項 $D1...D99$

$$Y = b0_A + b0_{D1}D1 + \dots + b0_{D99}D99 + b1_A gender + b1_{D1} gender \times D1 + \dots + b1_{D99} gender \times D99 + \epsilon$$

100 個迴歸式，合成一個迴歸式。有 1 個截距、99 個學校效果、1 個性別效果、99 個性別與學校交互作用，總共 200 個係數要計算。

以前對分組資料的處理

如果你已經受不了了.....

恭喜你，你已經準備好要學多層次分析了！

重新想一次

以分析 100 個學校 (A, B, C,) 為例，要瞭解性別對學生成績影響為例

- 分別做 100 迴歸，有 100 個截距，100 個斜率
- 想像 100 個截距形成一個分配，有截距的平均數 (固定效果部分)、截距的變異數 (隨機效果部分)
- 想像 100 個斜率形成一個分配，有斜率的平均數 (固定效果部分)、斜率的變異數 (隨機效果部分)

有截距平均、斜率平均 (固定效果)，截距變異數、斜率變異數，總共 4 個係數要計算。其實還有一個截距與斜率的相關。

重新想一次，還有更重要的

以分析 100 個學校 (A, B, C,) 為例，要瞭解性別對學生成績影響為例

- 如果要預測第 101 學校的男學生成績，前面有 200 個係數的式子 (固定效果) 怎麼用？
- 如果要預測第 101 學校的男學生成績，前面有 5 個係數的式子 (混合效果) 怎麼用？

固定效果、隨機效果

對於一個多層次資料

- 效果不依群體而變，即為固定效果
- 效果依群體而變，即為隨機效果
- 模型中兩個效果都有，就是混合效果模型
- 我們通常把隨機效果的平均數移出來當固定效果，因此隨機效果幾乎都伴隨固定效果
- 因此，稱隨機效果模型時，其實就是混合效果模型

示範

看一下固定效果模型與隨機效果模型 (Demo 1)

固定效果、隨機效果

- 固定效果模型
 - 兩百個係數分析 100 個學校的模型，是固定效果模型
 - 預測來自第 93 個學校的男學生成績，很準
 - 預測來自第 101 個學校的男學生成績，不能做
- 隨機（混合）效果模型
 - 五個係數分析 100 個學校的模型，是隨機效果模型
 - 預測來自第 93 個學校的男學生成績，沒很準
 - 預測來自第 101 個學校的男學生成績，可以做
- 效果選擇
 - 想知道的群體，資料都有，就設定為固定效果
 - 想知道的群體，資料沒有，就設定為隨機效果（與固定效果）

分析單位

以學生成績中性別的效果為例，學生巢套在班級內（其實又巢套到學校中）

- 學生性別（學生層級變項）
- 教師性別（班級層次變項）
- 班級學生性別比率（班級層次變項）

會不會女老師的班上，女生成績特別好？（學生性別 × 教師性別跨層次交互作用）

在沒有多層次分析前

50 班，每班 10 個學生；先只考慮學生性別與教師性別影響

- Aggregation
 - 以班級為分析單位 (50 筆)
 - 教師性別對學生平均成績影響
 - 沒辦法分析學生性別影響
- Disaggregation
 - 以學生為分析單位 (500 筆)
 - 學生性別、教師性別對學生成績影響
 - 同一位教師性別被複製 10 次，SE 低估
 - 教師性別更容易顯著，可是審查人不相信

生態謬誤

除了隨機效果、處理不同層次變項外，其實還有更重要的原因

- 定義

- 生態謬誤是指錯誤地將群體屬性的相關性推廣到個體層次的現象
- 生態謬誤通常發生在群體資料和個體資料之間的轉換和比較中

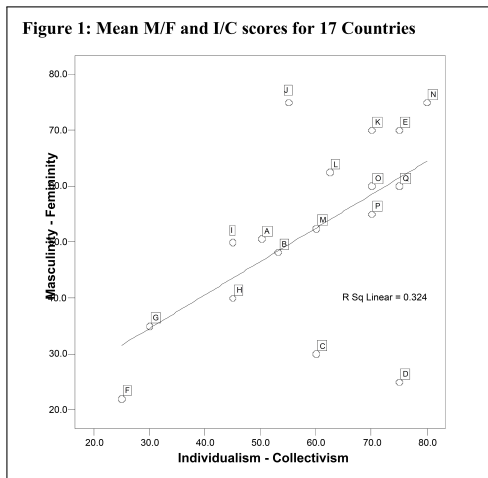
- 範例

- 研究發現某個地區的平均收入水平和健康狀況呈現正相關，但不能將這種關係推廣到每個個體身上
- 生態謬誤的發生原因：忽略群體內個體差異和異質性，忽略群體間差異和相互作用

- 影響

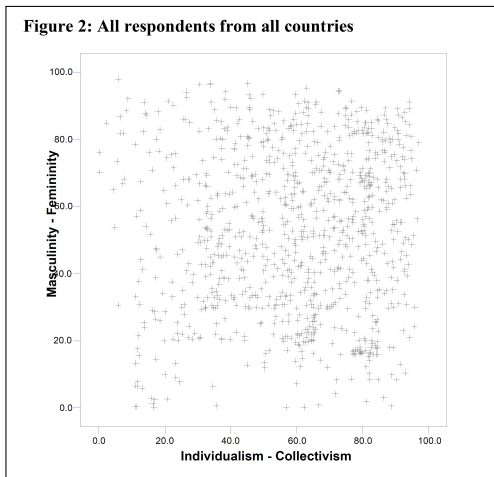
- 研究結果的失真和不準確性
- 政策制定和決策的錯誤和不當
- 社會公正和平等的忽略和破壞

生態謬誤的例子

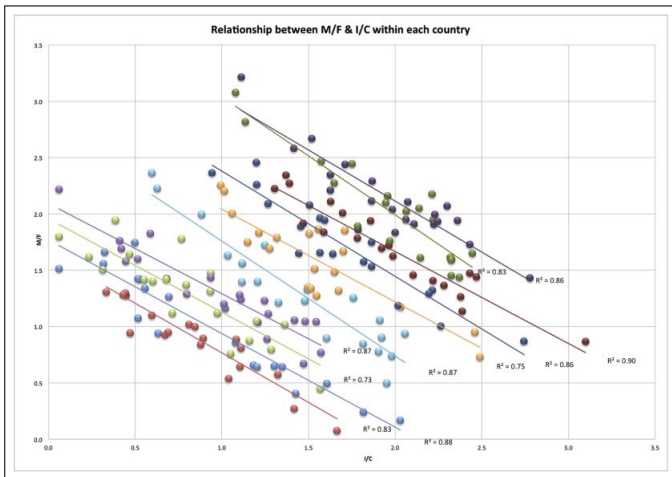


Winzar(2005)

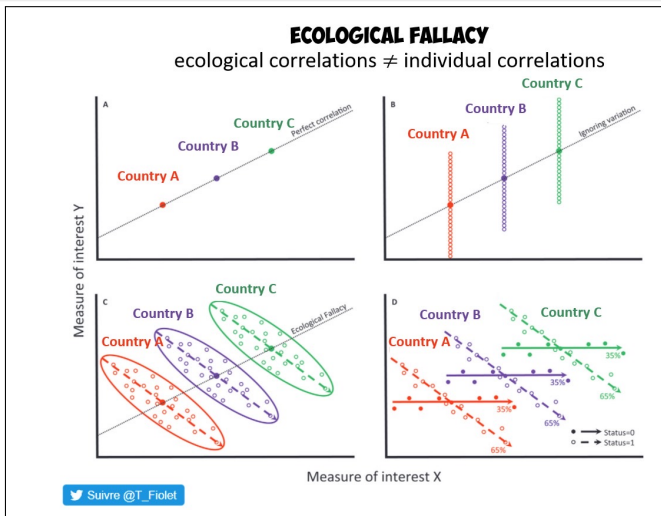
生態謬誤的例子



生態謬誤的例子



生態謬誤的例子



不考慮資料階層時，相關（關聯）其實混合了不同層次的訊息

分析時還是回到社會科學的根本

我們有興趣某個指標（常常對應分析單位的特性、被當成依變項）

- 1. 認識到分析單位在指標上有個別差異（依變項有變異）
- 2. 以其他變項解釋或說明分析單位的個別差異（依變項的變異以獨變項解釋）

以鄭老師夫妻的體重為例

再以薪水為例

我們有興趣不同人的薪水

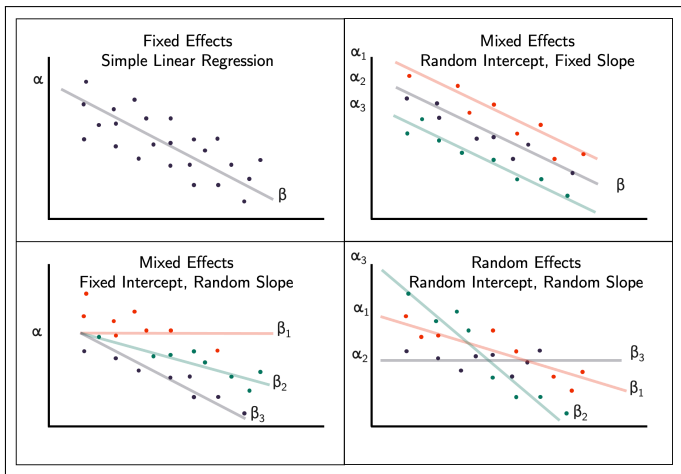
- 我們發現不同人（分析單位）的薪水不同（人.1）
- 我們可以利用變異，猜測某個人薪水的可能區間（人.1）
- 我們發現，不同公司、年資，人的薪水不一樣（人.2）
- 我們可以利用某人屬於某公司，猜測人的薪水區間（人.2, 公司.1）
- 我們可以利用某人所屬公司的特性，猜測人的薪水區間（人.2, 公司.2）
- 我們也可以利用某人年資，猜測人的薪水區間（人.2）
- 年資效果，又依公司不同而不同（人.2, 公司.1）
- 年資效果，依公司屬性不同而不同（人.2, 公司.2）

多層次迴歸模型：隨機截距、隨機斜率

- 多層次迴歸模型的基本形式
 - 單層次迴歸模型： $y = b_0 + b_1x + \epsilon$
 - 多層次迴歸模型： $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \epsilon_{ij}$
- 隨機截距模型
 - 定義和形式： $\beta_{0j} = \gamma_{00} + u_{0j}$
 - 解釋：隨機截距反映不同層次的群體差異
 - 應用：例如學校之間的學生成績差異
- 隨機斜率模型
 - 定義和形式： $\beta_{1j} = \gamma_{10} + u_{1j}$
 - 解釋：隨機斜率反映不同層次的斜率差異
 - 應用：例如 IQ 對學習成績的影響，隨學校而有差異

要把截距與斜率，想成 level 2 單元的性質

隨機截距與隨機斜率



示範

看一下隨機截距模型與隨機斜率模型 (Demo 2)

不同階層的獨變項：學校中的性別

- 資料階層
 - 學生層次：性別是學生層次的獨變項
 - 學校層次：性別比例是學校層次的群體特徵
- 獨變項的作用
 - 在學生層次，性別可能與學業成績、行為問題等有關
 - 在學校層次，性別比例可能與學校不同科的表現等有關
- 階層性模型
 - 隨機截距模型：探索學校之間性別差異對學業成績的影響
 - 隨機斜率模型：探索學校之間性別差異對學業成績和學生流失率的影響
 - 隨機截距與隨機斜率模型：同時探索學校和學生層次的性別效應

不同階層的獨變項：工作時間

工作時間對績效的影響

- 資料階層

- 員工層次：工作時間是學生層次的獨變項
- 組織層次：平均工作時間是組織層次的群體特徵

- 獨變項的作用

- 在員工層次，工作時間可能影響績效
- 在組織層次，平均工作時間可能與產業別有關，可能直接影響績效、也可能影響效率

工作時間的長短是相對的還是絕對的？

應該跟誰比？

當我們說，這家公司平均工作時間很長，我們在講公司的特性還是人的特性？

在平均工時不同的公司，工時長短對績效的影響會相同嗎？

不同階層的獨變項：對組織的忠誠度

這些變項，意味著什麼？

- 員工層次：個人對組織忠誠度
- 組織層次：組織內，個人對組織忠誠度的平均
- 組織層次：組織內，個人對組織忠誠度的變異

組織忠誠度平均數高，這表示組織的那個特性？

組織忠誠度變異數高，這表示組織的那個特性？

當變異數很高時，我們適合用「忠誠度平均數」表達組織的特性嗎？

我們可以把問題轉成信度的問題

- 把員工對組織的評價，想成一堆評分者對目標的評價
- 這類主觀的東西，應該評價要夠一致，才適合代表組織
- 有三個可以用的指標：ICC(1)、ICC(2)、 r_{WG}
- 概念上都是都是百分比 (0-1)，但因為計算關係，有時可能會小於 0

調節效果：What, Why & How

- What
 - 當 X 對 Y 的效果大小，受 Z 影響時，我們稱 Z 是 X 到 Y 的調節變項
 - 調節效果就是交互作用效果，數學上 X 、 Z 是對稱的，誰是調節變項就看切入點
- Why
 - 我們知道 X 影響到 Y ，但不希望只用增加 X 來增加 Y ，就需要找到調節變項
- How (如何檢驗調節)
 - 步驟一：檢驗調節效果 (交互作用項) 是否顯著
 - 步驟二：當調節效果顯著，我們會將 Z 固定在某些特定數值，畫出 X 對 Y 的影響

示範

看一下不同階層的獨變項與調節作用 (Demo 3)

中介效果：What, Why & How

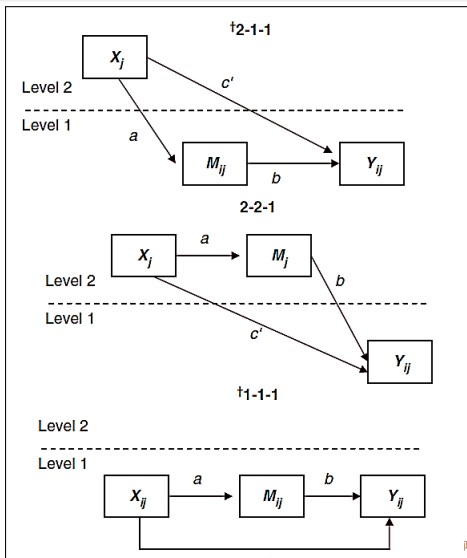
先談單一階層資料的中介

- What
 - 當 X 影響到 M, M 影響到 Y, 我們稱 M 是中介變項
- Why
 - 我們知道 X 影響到 Y, 但不清楚何以會這樣發生, 加入 M 會讓整個歷程可以理解
 - X 影響到 Y, 減少 Y 需要減少 X, 但 X 發生在過去或是不好更動, 我們得找到 M
- How (如何檢驗中介)
 - 間接推論中介：四步驟法 (Baron & Kenny 法)
 - 直接估計中介：估計中介效果 ($a*b$), 檢驗是否為 0 (Sobel test) 或是看信賴區間 (拔靴法)
 - 四步驟法的偵測力不足, 現在多用拔靴法

多階層資料中介效果

- Why
 - 在多層次迴歸中，Y 一定是 level 1，M 的層次一定低於或等於 X 的層次
 - X、M 不必然都是 level 1
 - 以二階層資料為例，可以有 1-1-1, 2-1-1, 2-2-1 的可能組合
- How (如何檢驗中介)
 - 間接法：Krull & MacKinnon (1999) 仿 BK 法，兩個式子
 - 直接法：估計中介效果 ($a*b$?), 檢驗是否為 0 或是看信賴區間 (拔靴法)
 - 2-1-1, 2-2-1, 中介效果都是 $a*b$ 對應之固定效果相乘
 - 1-1-1 時，中介效果的估計不再是 $a*b$ ，而是
$$E(a_j b_j) = E(a_j)E(b_j) + \sigma_{a_j b_j}$$

中介



示範

看一下 1-1-1 中介效果 (Demo 4)

行為科學中的時間資料

- 假設我們測量 100 位學生，由 1 年級到 6 年級的數學成績（經過垂直等化，可比較）
- 我們有 600 個資料點
- 過去處理方法：以年級為獨變項，平均成績為依變項
- 只剩 6 個資料點，個別差異全部不見了
- 會不會有生態謬誤？

個別成長模型 (Individual Growth Model)

以混合模型方式重想

- 想像 100 位學生，各有一條迴歸線
- 100 條迴歸線，會有 100 個截距與 100 個斜率
- 截距是起點，斜率是學習效率
- 可以將個別成長模型，想成是把時間資料，抽出起點與學習效率等各人特徵的方式
- 把起點與學習效率集群分析，會不會有不同族群？起點低又慢、起點低但快、起點高但慢、起點高又快？
- 可以進一步問，哪些變項影響起點？哪些變項影響學習效率？

個別成長模型 (Individual Growth Model)

- 模型設計

- 說明個體的特質 (如學術成就測試分數或身體測量等) 隨著時間推移的變化情形
- 目的是描述個體的變化，並評估個體因素對於變化的影響，例如：隨著時間推移，是否有一個初始水平、斜率、加速度等

- 應用範圍

- 教育研究：探討學生的學業成就、學習動機等隨著時間的變化
- 心理學研究：探討心理健康、憂鬱隨著時間的變化
- 醫學研究：探討身體健康狀況隨著時間的變化

要注意一下時間的原點、此外可考慮二次式或多次式

示範

看一下成長模型 (Demo 5)

在實驗資料中考慮隨機效果

我們設想一個單因子（三個水準）的受試者間設計

- 給定一個水準，假設有 30 個參與者，並需要對 10 個刺激作反應
- 反應假設是反應時間（連續資料）
- 一個嘗試對應到三個座標：獨變項水準、參與者編號、刺激編號
- 獨變項可能是固定或隨機，參與者是隨機
- 這是多階層資料（trial 是 level 1, 參與者是 level 2, 變項是 level 3)

回想一下實驗設計（Kirk!），獨變項是固定或隨機，會影響對應的誤差項

在實驗資料中考慮隨機效果

其實更複雜

- 一個嘗試對應到三個座標：獨變項水準、參與者編號、刺激編號
- 刺激其實也常常是隨機效果
- 當我們要分析獨變項的效果，針對每個水準，其實對應到方形（而非長條）的資料（ 30×10 ）
- 我們可以有 by subject 的分析，把 10 個刺激的 RT 平均，得到 3×30 個反應 (??)
- 我們可以有 by item 的分析，把 30 個參與者的 RT 平均，得到 3×10 個反應

在實驗資料中考慮隨機效果

「行為」科學？還是「平均行為」科學？

- 我們可以分析每一個 trial 的 RT，這時 trial 同時巢套在參與者與刺激內
- 參與者與刺激都是隨機效果，這是 cross random effects，不是 nested random effects
- lme4 寫起來很容易，算起來很難（不過我們不在意）

效果量？

以混合效果模型處理心理學實驗資料

- 可以同時處理刺激與參與者的隨機效果，不用區分 by subject 或 by item
- 不用再變項是隨機或固定時，如何設定誤差項，混合模型會搞定
- 混合效果擅長處理巢套關係，可以順利處理 nested design
- 行為科學，而不是平均行為科學，效果量或預測值才對

示範

看一下實驗資料的分析 (Demo 6)

常見問題

- ICC(1), ICC(2), r_{wg} 應該要多高？
- 一定要 centering 嗎？對總平均，還是群體平均？
- 需要多少樣本數？第一層與第二層分別需要多少？

常見問題-1

ICC(1), ICC(2), r_{wg} 應該要多高?

指標	切點	使用篇數
r_{wg}	0.7	12
ICC(1)	0.05	1
	0.1	0
	0.12	4
ICC(2)	0.6	4
	0.7	7

朱海騰 (2020) 針對 2014-2019 篇 JAP 文章的統計，83.53% 報告 r_{wg} ，91.76% 報告 ICC(1)，85.88% 報告 ICC(2)。建議三者都報告。

常見問題-2

一定要 centering 嗎？對總平均，還是群體平均？

- 截距是 $X = 0$ 時的依變項期望值，由於我們假設隨機截距是常態， X 的原點可能需要適當訂定
- 對總平均置中，相較於以原始分數，其實不會改變檢定結果，只是讓軟體更好計算
- 以原始分數估計層次一變項的效果，可能同時包括群體的效果與個人效果，對群體平均置中，可以區分這兩者

常見問題-3

需要多少樣本數？第一層與第二層分別需要多少？

- 樣本數的問題，牽涉到估計精確性與檢定力
- 通常建議每群體最好有五個（有人建議三個）
- 如果要有.8 的檢定力以偵測中等以上的固定效果，當隨機效果中等時，建議 $L1 > 30$ ， $L2 > 200$ (Arend & Schäfer, 2019)。
- 檢定力與效果量、ICC 有關，可以找到給兩層次資料的檢定力分析軟體軟體 (PINT)。

Thanks!

謝謝!