

## 社會四 林煜翔

### 文章摘要

統計學上的多變量分析是適合農業科技的研究方法，原因在於在農業相關的問題中，研究者感興趣的問題原因通常不是單一來源的，而是受多個變項影響，因此多變量分析中著重探討變項與變項間的交互關係會是適合農業科技，或其他著重多元成因的科學，例如社會、心理、經濟的學科的研究方法。

作者在本文中介紹 7 種主要多變量分析方法背後的概念以及案例探討：

#### 主成份分析：

主成份分析的主要概念在於將多個變數產生成數個彼此獨立的綜合指標，而這些指標在數學上的意義是能夠解釋最多變異為前提。使用此方法的原因在於在多變量分析中，變項與變項間往往是彼此相依(dependence)的，然而該相依性違反回歸分析的假設，也就是變項與變項間應該是線性獨立，而主成份分析則是將不同變數線性組合成數個彼此獨立的變項，以期估計結果將能達到不偏。作者提到主成份分析不僅能夠將數個變項縮減成較少個指標，另一功用在於能篩選變異貢獻大的指標，捨棄變異貢獻少的指標已達到第二次降低維度的效果。

#### 因素分析：

因素分析的使用前提同樣是因為多變量分析中變項與變項間彼此並不獨立，因素分析的概念在於將原始的多個變項以較少的因素表示，也就是說原始變項可以寫成較少因素的線性組合，透過該方法達到降低維度的效果。作者特別提醒許多研究會將因素分析與主成份分析混為一談，然而若思考主成分及因素如何產生，便可理解兩者差異，即主成份是多個變項的線性組合，因素則是線性組合後得以表示為原始變項，也可以說成主成分是變項的指標，而原始變項則是因素的綜合指標。

### 對應分析：

對應分析使用兩個主成份作為二維的座標軸，圖上會有原始變數及樣本的散佈位置，透過樣本間位於座標圖上的距離判斷彼此的性質。因此如作者前述，適合統計分析前座描述性統計，透過清楚的方式展示不同變項、樣本間的異同。

### 集群分析：

作者提出過去分類的問題經常仰賴經驗或是專家的知識，缺乏數值方法，而集群分析則是以數學的方式，依據不同變項將未經分群的樣本進行分類，以演算法的方式將樣本分為不同群組，其最主要的目標在於分在同群的樣本彼此的變異要小，而不同群的樣本間彼此變異要大。然而分群的數目最終仍須仰賴個人的專業判斷，也就是需要回到理論上的基礎，或是將分群的結果視覺化，觀察最好的分群數目。

### 判別分析：

判別分析的功能同樣是分類，然而和集群分析的差異在於判別分析已經有既定的分類，並透過選出可以代表不同分類的樣本產生判別函數，之後將未知分類的樣本帶入，即可得到分類結果。

### 路徑分析：

路徑分析透過因果分析模式說明研究假設中的因果關係，其作法流程為研究者根據研究者所假設的因果關係圖，針對各個不同的變項進行回歸分析，接著得到變項間的回歸係數，並篩選回歸分析結果中顯著的回歸係數，並可得到路徑分析的結果。

### 典型分析：

典型分析主要的功用是判別兩組變項的關係為何，也就是超越前面多個自變項對單個依變項的架構，而是多個自變項對多個依變項的分析。具體的做法是透過主成份分析的方式分別產生出可以寫成兩組多個變項線性組合的主成分，且

這兩個組成分共變異為最大。

## 應用案例：

### 主成份分析：

有時候問卷的設計不見得會有題組與題目之分，也就是說研究者並沒有假設任何題目可以歸到特定的範疇，在此狀況下就可以利用主成份分析進行維度縮減，並在事後針對各個主成份進行解釋。例如若要使用問卷的範式判斷一個人是否有憂鬱症時，題目可能會設置數十個情境題讓受試者夠選，而研究者可以透過主成份分析將原始變數降維至若干個主成份，其中可能就會包含情緒管理、人際互動、家庭關係等等意義的主成份。

### 因素分析：

我做過的因素分析是關於行銷的研究，針對每一次訂單中的商品產生商品與商品之間的相關係數矩陣，進行因素分析後觀察不同商品在各個因子的係數，並篩選係數大於某值閾，最後針對個因素做解釋，找到容易被共同買單的商品，也就是通稱的“一籃子”分析法。

### 對應分析：

延續上述主成份分析的案例，對應分析進一步將主成分的結果視覺化為二維的圖形，假設上述使用問卷調查憂鬱症的案例最後得到兩個主成分，則可以透過對應分析觀察原始變項與兩個主成份的關係為何，也可以觀察受試者們彼此的差異性，若進一步將受試者本身的其他屬性，例如性別、年齡等變項透過改變散佈圖中點的形狀、大小、顏色進行變化，則可以在投入研究前得到初步的假設。

### 集群分析：

如果要做社會階級的分類，可能需要考量經濟資本、文化資本、社會資本、教

育資本等概念，假設有收入、藝文支出（作為代表文化資本的變項）、個人網絡規模、教育程度等變項，即可透過集群分析的方式產生不同的群體分類。

### 判別分析：

我曾經做過將遊戲評論分類的專題屬於判別分析。我們先透過人工的方式將遊戲評論分成各個類別，接著從評論中產生有意義的變項，例如動詞數、形容詞數、評論的情緒分數等，並建構預測的模型，當時用的是 **multinomial logit**，最後將新評論丟入得到預測結果。

### 路徑分析：

假設要判別教育程度對於收入的影響，研究通常會放入父母教育程度、父母收入、子女教育程度、子女收入等變項，由路徑分析架構即可說明研究者對於上述四個變項的因果模式假設。

### 典型分析：

典型分析我想到的例子是疾病相關的研究，因為疾病通常是伴隨併發症的，且導致疾病風險因子也是多樣的，所以典型分析應該能夠找出疾病與風險因子的關係。