

學習單 3- 集群分析

金企四甲 林嘉柔 409411167

- 從 SRDA 網站上下載 教育相關的資料來做集群分析：

Q1 針對此數據在 R 軟體執行階層式集群的 AGNES 演算法、DIANA 演算法來作分群，並說明與解釋其分析的結果。

Q2 針對此數據在 R 軟體執行分割式集群的 K-Means 演算法、K-Medoids 演算法來作分群，並說明與解釋其分析的結果

- **資料主題：**「實習學生的網路社會資本、求助行為、自我調整學習與教師信念的關聯之研究」。

一、問題描述：

對於進行教育實習的學生來說，雖然學校安排了輔導教師給他們並提供完善的輔導制度，但卻仍然面臨著在陌生環境中與輔導教師互動、班級經營、溝通困難和人際關係等問題。而先前研究指出，他們即便尋求輔導教師或其他人員的協助，也未必能得到充分的解決，因此，為因應這些困境，實習學生往往會透過網路社群及其他社會資本來尋求幫助和慰藉。

尋求幫助後，實習學生需要透過自我調整，根據不同的問題情境與回應調整實習任務，以因應教育專業的需要。而且在這些過程中，除了應用所學知識與技能外，他們也可能對教師的工作有重新地瞭解。其藉由不斷地與人互動、解決內在衝突並調整對此工作的看法，若其新的觀點能獲得他人認同，便可能塑造正向的教師信念。

然而，現今對於實習學生如何運用網路社會資本與此求助行為如何影響他們的自我調整學習策略，以及這些對教師信念的影響尚缺乏深入探討。因此，此教育資料主要透過深入了解他們在教育實習期間的行為和思考來進行研究，希望為提升教師培訓過程中的支援措施提供資訊。

參考資料：[檔案連結](#)（研究動機與目的）

→→ 而此份學習單主要透過階層式集群的 AGNES 演算法、DIANA 演算法以及分割式集群的 K-Means 演算法、K-Medoids 演算法來作分群，依據資料之間共同屬性，將較為相似、相近的資料聚集在一起，並說明與解釋其分析的結果。

二、數據分析：

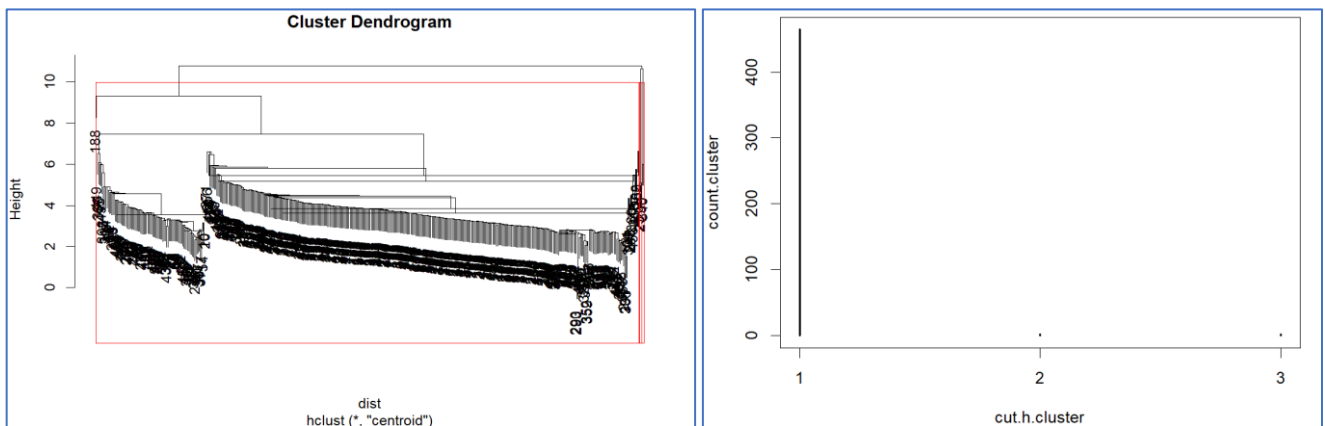
● 資料說明與處理

變數多為五點量表 1~5 數值，代表非常不同意至非常同意，而資料報告中已透過信效度分析得知題項 c06、d14、e02 的相關性不顯著（因 $r < 0.05$ ，採 0.95 的信心水準與 5% 的信賴區間）因此先將此 3 題從資料中刪除。也在單因子變異數分析中得出無論實習生參與哪一個屬性學校的教育實習，在探討變項上沒有存在差異性，因此也將基本資料中觀察實習學校屬性的題項 school 刪除。

此外，題項 b0101~b0104 為將勾選「尋求協助時會聯繫的對象」變成類別型變數（勾選為 1，無勾選為 0），為了一併探討「聯繫對象」的重要性與對教師信念的關聯，以及方便進行整體的數據分析，因此將 0、1 中勾選家人的 1 改為 1、大學老師的 1 改為 2、學長姐或同學（未在同一學校實習）為 3、朋友為 4，而因為題項可複選，所以改取平均值歸類至同一個變數 b01。（先以數值大小概略分析，看聯絡對象是否會有顯著影響，如果此項有顯著影響也可以再重新區分開來做為未來研究的方向之一。）

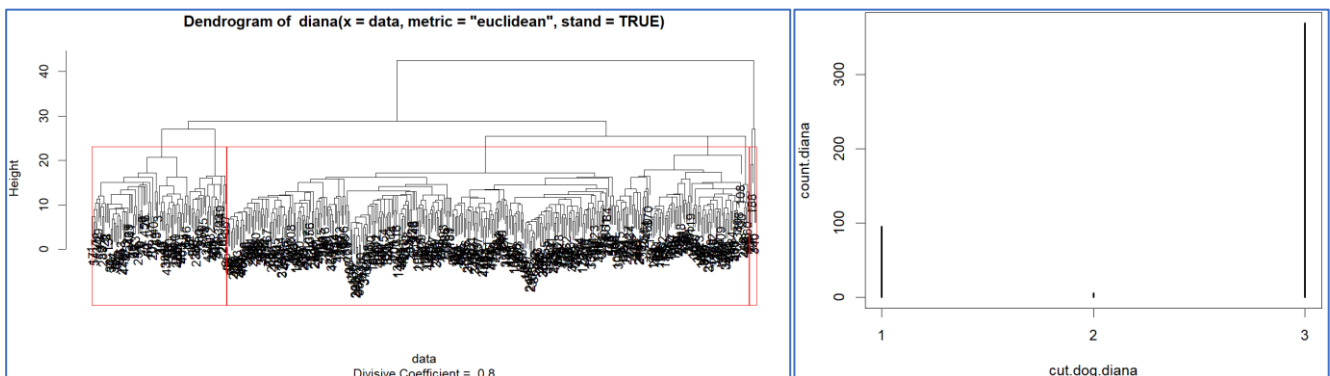
1. 階層式集群 - AGNES 演算法

左圖為樹狀圖並以紅框分成 3 群；右圖為每群個數之長條圖。



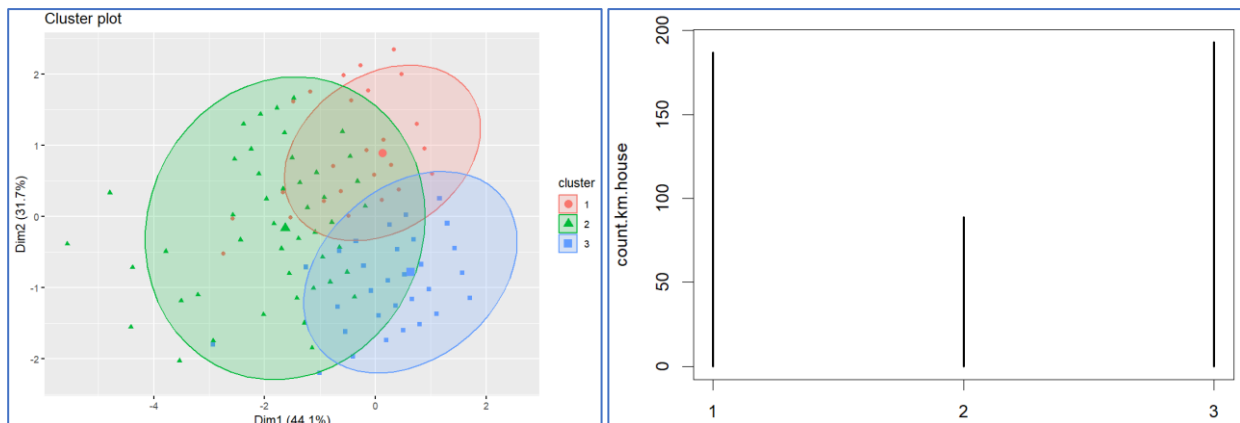
2. 階層式集群 - DIANA 演算法

左圖為樹狀圖並以紅框分成 3 群；右圖為每群個數之長條圖。



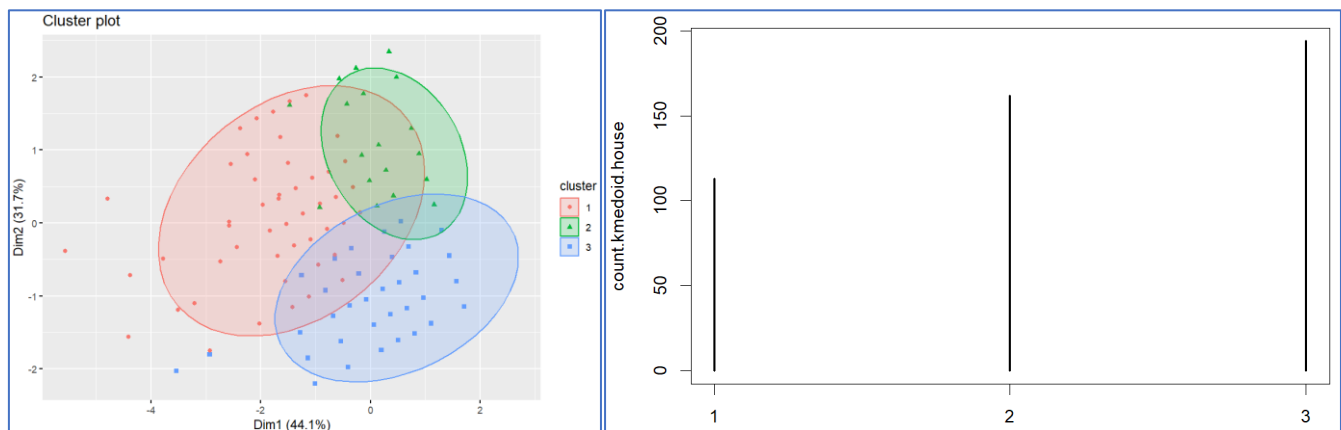
3. 分割式集群 – K-Means 演算法

左圖為分群圖；右圖為每群個數之長條圖。



4. 分割式集群 – K-Medoids 演算法

左圖為分群圖；右圖為每群個數之長條圖。



從以上四種演算法的結果來看，整體來說：

- 階層式集群的結果中絕大多數之樣本資料皆被分至同一群，而分割式集群的結果中樣本資料的分類較為平均，由此評斷以分割式集群的分群方式較佳，較能區分樣本資料之間之共同屬性與差異。
- 而在分割式集群中，又以 K-Medoids 演算法的分群結果較佳，數量分佈較平均之外，群體間的重疊部分也較少，相對來說較能區分樣本資料之間之共同屬性與差異。
- 最後，整體來說其實分群的结果看起來都沒有辦法分得很乾淨，可能是樣本資料間原本差異就不算很大的關係，所以才有這些重疊的部分；不過未來研究中若是能以 K-Medoids 演算法的分群結果來深入探討，或許也能得出不錯的結論，進而幫助提升教師培訓過程中的支援措施。