

# 107 學年第一學期高維度資料分析期末報告規定

吳漢銘  
國立臺北大學統計學系

December 23, 2018

## 1 規定

1. 分組: 每組一至兩人, 最多不得超過三人。
2. 以本學期學到的方法, 分析資料, 並簡報結果。(見下節)
3. 請勿抄襲或從網路上 Copy, 請認真!
4. 要上傳的文件: (1) 簡報檔。(2) 影片檔。(3) R 專案 (壓縮) 檔。
5. 截止日: 2019/01/08(二); 24:00。(請提早上傳, 見底下上傳方法)
6. 老師評分 (等級給分): 書面報告、上台表現、統計方法適用度。
7. 自評、每組互評 (等級給分): 上台表現、流暢度、表達能力、簡報製作。
8. 第 18 週上課, 播放各組影片, 同學們需自評、每組互評 (等級給分): 上台表現、流暢度、表達能力、簡報製作、影片製作等等。

## 2 組別/上傳目錄 (注意: 皆沒有空格)

上傳 FTP: 120.126.135.22」, 帳號: hdda。密碼: xxx。

01-陸薇安

02-劉又瑄-蔡岱融-藍雍淵

03-駱哲宇-黃昱傑-蔣竹軒

04-王姿文-康益豪

05-葉憶雯-陳涵傑

06-鄭琪珍-張維仁

07-傅國璋-武靖祐

08-鄧鈞洲-許嘉銘

09-沈冠青-胡慧禪

10-陳柏鈞-張一凡

11-薛愛蓉-顏振庭

12-蕭子修-許巍瀚

13-葉家齊-許哲瑋

14-許馥欣-陳寶寧

### 3 資料分析

1. 資料來源: 教學網站所列, 或自行尋找。
2. 資料維度大小: 所使用的資料 (若名稱為 DataSetA) 為  $n \times p$ , 其維度大小最好是傳統的  $n \gg p$ 。將此資料隨機抽樣, 得到兩資料集 (名稱為 DataSetB、DataSetC), 其中 DataSetB 維度大小為  $n \approx p$ , DataSetC 維度大小為  $n \ll p$ 。
3. 資料處理: 檢查資料、遺失值、轉換。
4. 統計方法: 敘述統計、探索性資料分析、維度縮減方法 (至少需包含 PCA 及 ISOMAP, 要有驗證)、群集分析 (要有 heatmap、要有驗證)。
5. 若統計方法有不適用的之處, 請做適當之處理。
6. 要比較三個資料在各統計方法下分析結果的異同。最好以圖表呈現。

### 4 R 專案檔

1. R 專案, 請命名為「HDDA-組別-R」, 例如: HDDA-01-R。
2. 專案中主程式檔請命名為 main.R。(教師只要 `source("main.R")` 就可執行程式, 並得到全部結果。)
3. 請上傳壓縮檔 (zip, rar)。例如: HDDA-01-R.zip。
4. 專案內容需包含所有的 R 程式碼、或圖片檔、或資料檔等等。(資料檔請放在專案下之 data 子目錄)
5. 如何建立一 R 專案檔, 請見 [http://www.hmwu.idv.tw/web/R/A01-hmwu\\_R-Introduction.pdf](http://www.hmwu.idv.tw/web/R/A01-hmwu_R-Introduction.pdf), 第 35/53 頁。

## 5 簡報檔

1. 檔名: 組別-姓名-Slides.pptx, 例如: 01-陸薇安-Slides.pptx。
2. 上傳檔案格式: PowerPoint (ppt, pptx); 可用其它簡報軟體 (pdf)。
3. 叮嚀: 一份資料分析簡報不需要做多, 但要做完整。
4. 封面: 主題、組員學號名字、單位、日期。
5. 導論:
  - (a) 分析 (研究) 背景、動機、目的。
  - (b) 想了解的問題是什麼: 列出問題由大方向至小細節, 先不管問題是否可解決。不能解決的原因為何? 可放至報告的後面討論。
  - (c) 定義問題, 其相對應的統計問題是什麼?
  - (d) 文獻回顧: 這些題目, 有誰做過類似的? 結果如何?
  - (e) 分析此資料需要什麼「背景知識」domain knowledge。
  - (f) 整個研究 (資料分析) 有什麼重要性?
  - (g) 整個研究 (資料分析) 的困難點是什麼? 如何解決?
  - (h) 分析步驟: 流程圖。
6. 資料來源及使用要明確:
  - (a) 如果是問卷調查資料, 就要有「實際問卷」於附錄。
  - (b) 所收集的資料的格式為何? 本分析採用哪些資料 (觀察值、變數、欄位)? 為何?
  - (c) 是否有額外的輔助資料? (例如: 經緯度)
  - (d) 收集資料是否足以回答問題? 是否有其它可補充的資料?
7. 資料處理:
  - (a) 資料變數 (連續、類別變數) 摘要表: 編碼、值域範圍、五數綜合、遺失值比例、次數。
  - (b) 探索資料 (EDA): 資料的分佈為何? 單變數圖形、雙變數圖形、多變數圖形、(列聯) 表格。
  - (c) 資料是否有群性、有特別樣態、趨勢?
  - (d) 遺失值、離群值如何處理?、需要做轉換嗎?。資料的處理方式是否合理?
8. 資料分析:
  - (a) 所使用的 (統計) 方法為何? 其假設為何?
  - (b) 評估統計方法的適用性 (假設、限制、優缺點)。

- (c) 多個統計方法的比較。
- (d) 模型配適、模型選擇、結果解釋及原因。(需解釋分析的結果)

9. 結論與討論:

- (a) 本主題的結果、討論、結論、貢獻、價值是什麼。
- (b) 是否有新想法，有無其它的問題可提出? (以不同角度去問問題: 分析者、資料提供者、社會大眾等等)
- (c) 本資料分析的限制、未解決的部份、未來可再改進或延伸的部份是什麼。

10. 其它:

- (a) 誌謝詞: 寫下要感謝的人事物吧! 或做完本主題的心得 (最難的部份是什麼? 我的收穫是什麼?)。
- (b) 參考書目: 列出參考書目, 網址等等。
- (c) 附錄: 列出所使用的 R 套件及其指令: `command {package}`。
- (d) 附錄: 條列出每個組員在本專案中負責的工作。
- (e) 附錄: 每個組員列出他本學期希望得到的分數。

## 6 影片錄製

1. 每位組員都要 (上台) 簡報。簡報一人影片 8 分鐘、簡報二人影片 10 分鐘、簡報三人影片 15 分鐘。
2. 影片最後可放個人心得、感想、NG 鏡頭 (不計入簡報影片時間)。
3. 組員一人, 可在個人電腦螢幕前簡報並錄製。組員二人以上, 需有投影機畫面 (畫面固定不要晃動), 注意燈光 (不可太暗)、角度、簡報者的音量。
4. 每組報告的影片, 會放在教師的教學網站上, 供日後同學們參考。

5. 錄影

- (a) 手機錄影 (mp4 格式)
- (b) 畫面大小 1280x720 (或至少 1080p)
- (c) 影片編輯軟體 (以下供參考; 用別的軟體也可以)  
Avidemux is a free video editor (影片剪接): <http://fixounet.free.fr/avidemux/>  
Windows Movie Maker 2016 (影片/圖片/配樂/字幕/過場效果): <http://www.windows-movie-maker.org/>

6. 歷年學長姐上台報告的影片 (注意同學們報告時的優缺點)

- (a) 105-2-電腦概論: <http://www.hmwu.idv.tw/index.php/teaching/265>
- (b) 105-2-巨量資料分析: <http://www.hmwu.idv.tw/index.php/teaching/267>
- (c) 106-2-高維度資料分析: <http://www.hmwu.idv.tw/web/teaching/106-1/hdda/final-video/>

## 7 上台簡報要點

1. 開場、結束，要跟評委和聽眾問好。
2. 時間控制很重要，所以正式上場前，應自行練習。不順暢/不熟練的報告過程是浪費大家的時間。
3. 動機最好用一個 (生活上的) 故事陳述。
4. 開場三句話內就應把主題清楚表達。
5. 主題要明確指出所分析的問題，不要太籠統/花俏/浮誇。
6. 要把簡報說成一個完整吸引人的故事。
7. 時間有限，要開門見山。陳述跟主題直接相關的資訊。
8. 正式上場前，應自行練習並「錄影」，事後改善。
9. 請使用 PowerPoint 等簡報軟體報告 (需有頁碼)，勿直接開放 word 講解。
10. 簡報是要講給別人了解，而不是自己講自己懂就好。
11. 簡報內容，懂的才放，不懂的不要放。
12. 要有一頁大綱。
13. 報告時，不要一直看電腦，不要一直看著投影布幕，要看一下聽眾。
14. 眼光看評委觀眾，不是一直盯著投影布幕或電腦。
15. 若有錄影，應注意畫面要固定，不可晃動。主體 (投影畫面和簡報人) 要清楚不可太小，鏡頭不要拉太遠。
16. 簡報時，最忌一直看著投影片唸，而忘了與觀眾互動。
17. Powerpoint 的 (佈景主題) 顏色搭配，應以觀眾的觀看舒適度為考量。尤其是字體與背景的顏色不可太相近。
18. 字體要大，句子要短。
19. 投影片裡的流程，或箭頭，可參閱「商業簡報」裡面用的圖示。
20. 小技巧: 設計互動問答。放 (短) 影片、音樂。
21. 有需要簡報筆或其它輔助工具，應事先準備，且要事先勘查場地，確定設備。
22. 組員如何輪流上場、一人簡報另一人操作電腦要有默契，用簡報筆，都要排練過。
23. 要展現團隊合作。站好，身體不要一直晃動。報告時，重覆之語助詞不要太多。

24. 引用圖片/表格/影片，要註明來源。
25. 請說專業用語。(e.g., 佔「幾 pa」=> 佔「百分之幾」)
26. 最後一頁，不要只放「Thanks」或「The END」，要放「Acknowledgment」，感謝相關的人事物。最好是一個問句! 給聽眾反思的機會。
27. 報告流程及重點: 問候、分析動機/目的、資料介紹、定義問題、資料處理、EDA、分析方法、評估、結果/結論、貢獻、建議、感謝/結束語。
28. 要何如回應提問者的問題?
29. 請查詢並學習「如何做簡報」等等相關書籍。請查詢如何上台報告: google/博客來/圖書館/youtube
30. 要大方、有信心、不要怕、不要笑場，眼睛看觀眾跟評審。(全部組員都是)
31. 上台報告，穿著整齊乾淨即可，上衣需要領子，不可著短褲或拖鞋/涼鞋。
32. 保持熱情。