

日期: 2024/06/18(二) 期末考: Open Book
授課教師: 吳漢銘 (政治大學統計學系副教授)

請仔細閱讀每一個注意事項 (禁止討論)

1. 考試期間 (第一階段: 06/18 (二) 13:10-15:50, 於上課教室。)
 - (a) 可參考課本、上課講義 (包含電子檔) 及其它資料, 但不能與別人討論。
 - (b) 可使用計算機、自己的筆記型電腦、平板電腦和手機, 但不能使用通訊 APP。
 - (c) 全程可上網查詢, 但不能用通訊軟體 (例如: FB/LINE) 討論, 也不可抄襲網路上之程式碼, 不可使用 chatGPT 之類的 AI 工具。一定要親自做。
 - (d) 有問題者, 請舉手發問。勿與同學交談。不按照規定作答者, 酌量扣分。
2. 下載題目卷: 於教學網站 (<https://hmwu.idv.tw>) 下載題目卷 (zip 檔), 請使用所提供的樣版檔做答: 「學號-Consulting-FinalExam.tex」, 「學號」改成自己學號。
3. 上傳答題檔案:
 - (a) 於教學網站首頁登入 [作業考試上傳區], 帳號: consulting112。密碼: xxx (上課教室號碼)。
 - (b) 請上傳「學號-Consulting-FinalExam.tex、學號-Consulting-FinalExam.pdf、學號-Consulting-FinalExam.R」。 (資料夾: 「20240618_第 1 階段_FinalExam」。) 上傳完畢, 請登出。
4. 答題檔案原則:
 - (a) 作答時請隨時存檔, 檔案毀損、變空白或消失, 請自行負責。
 - (b) 若上傳檔案格式錯誤, 內容亂碼, 空檔等等問題。請自行負責。
 - (c) 依公平性原則, 檔案不得上傳兩次以上。若上傳兩次 (含) 以上 (以第一次上傳檔案計分)、格式不合等等酌量扣分。
 - (d) 如果上傳網站出現「空白頁», 請將滑鼠移至「網址列」後, 按「Enter」即可。若再不行, 請換 (Edge/Firefox/Chrome)。
5. 上傳完畢, 完成考試, 即可離場。
6. 第二階段 (自由選擇是否要做答): 至 06/19 (三) 24:00 止, Take Home。
 - (a) 同一份考題做答再上傳 (資料夾「20240619_第 2 階段_FinalExam」)。(註: 不是只做第一階段沒做的題目, 要全做。)
 - (b) Open Book, 可使用網路查尋。不可問人, 不可使用 chatGPT 等 AI 工具, 一定要親自做。
 - (c) 最後得分 (註: 各階段分數若超過 100, 以 100 計再計算最後得分)。
 - i. 最後得分 = 第一階段。
 - ii. 最後得分 = 第一階段 \times 0.6 + 第二階段 \times 0.4。

我已經仔細閱讀上述各注意事項, 若有違背, 會自行負責。

日期: 2024/06/18(二) 期末考: Open Book
授課教師: 吳漢銘 (政治大學統計學系副教授)

1 探索式資料析 (40%)

資料

- 資料名稱: 家庭收支調查-家庭消費支出結構按消費型態分
- 資料來源: 政府資料開放平台
- 資料說明: <https://data.gov.tw/dataset/6588>
- 資料檔: 005-家庭消費支出結構按消費型態分.csv
- 資料檔原網址: <https://shorturl.at/gTJuT>

問題

(50%) 以上述給定的資料進行探索式資料分析。竭您所能及所學 (課堂上或自學), 以統計量、統計圖型及統計方法探索此資料, 並給予每個步驟或做法一個簡單的說明和解釋, 而不是僅列出程式碼和結果。請自行以 `\item[1.x]` 編號列舉每一個步驟或做法, 例如 1.1 敘述統計、1.2 索引圖、1.3 直方圖、...。依此資料探索之結果, 最後列出一些想了解或要解決的問題。這些問題需要其它的輔助資料嗎? 若有請說明。(註: 圖之標題, xy 軸標號及圖例說明 (legend) 需完整)(採用分析的項目和方法愈多, 得分則愈高)

1.1 敘述統計: {一些簡單的說明和解釋}

1.2 索引圖: {一些簡單的說明和解釋}

1.3 直方圖: {一些簡單的說明和解釋}

...

1.9 想了解或要解決的問題: {列舉}

- ...

- ...

- ...

2 維度縮減 (30%)

資料

有三類不同品種的小麥種子: Kama, Rosa 和 Canadian。每一品種小麥皆隨機選擇出 70 顆麥粒，接受檢測。每顆麥粒量測以下 7 種幾何屬性: (1) 面積 (area, A); (2) 周長 (perimeter, P); (3) 緊密度 (compactness $C = 4\pi A/P^2$); (4) 麥粒長度 (length of kernel); (5) 麥粒寬度 (width of kernel); (6) 不對稱係數 (asymmetry coefficient); (7) 麥核溝長度 (length of kernel groove)。

- 資料名稱: 小麥種子 (seeds Data Set)
- 資料來源: UC Irvine Machine Learning Repository
- 資料說明: <https://archive.ics.uci.edu/ml/datasets/seeds>
- 資料檔: `seeds_dataset.txt`
- 資料檔原網址: <https://archive.ics.uci.edu/static/public/236/seeds.zip>

問題

以上述給定之資料，應用三種維度縮減方法: Principal Components Analysis (PCA) · Multidimensional Scaling (MDS) · Isometric Mapping (ISOMAP)。(參數自選)(註: 圖之標題， xy 軸標號及圖例說明 (legend) 需完整)

- 2.1 (5%) 讀取資料，印出資料摘要。
- 2.2 (5%) 以原始給定資料，畫出降維後的二維資料散佈圖，每一觀察值需以顏色標上所屬小麥品種。(三張圖一頁)
- 2.3 (5%) 將資料的每個變數標準化，畫出降維後的二維資料散佈圖，每一觀察值需以顏色標上所屬小麥品種。(三張圖一頁)
- 2.4 (5%) 將資料的每個變數常規化 (即常規化後變數值的範圍為 $[0, 1]$)，畫出降維後的二維資料散佈圖，每一觀察值需以顏色標上所屬小麥品種。(三張圖一頁)
- 2.5 (5%) 承 2.3，畫出 Co-Ranking Matrix。(三張圖一頁)
- 2.6 (5%) 承 2.3，印出 LCMC 值 (參數自選)。那個維度縮減方法較適合此資料?

3 問卷分析: 題項分析與因素分析 (50%)

資料

- 資料名稱: 幸福溫度計調查: 2019 年第一波調查全會員樣本
- 資料來源: 中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫 (SRDA)
- 資料使用說明: C00391.pdf。調查執行報告: C00391r.pdf。網路問卷: C00391q.pdf
- 資料檔: C00391_data.csv
- 資料檔網址 (需登入會員): https://srda.sinica.edu.tw/datasearch_detail.php?id=3529

問題

以「< 幸福感指標 >」分量表 V10 ~ V17 為分析資料。

- 3.1 (10%) 讀取資料，以 heatmap 呈現資料行列排序前與排序後之結構。(列為受試者，行為題項。列之相關以歐氏距離計算，行之相關之 $(1 - \text{相關係數})$ 計算)。
- 3.2 (20%) 進行**題項分析** (自行編號及標題，列舉每一做法或步驟，需有一些簡單說明)(參數自選)。綜合上述之分析，哪些題項是需要刪除的? (請寫刪除的理由。)(**採用分析的項目和方法愈多，得分則愈高**)
 - 3.2.1 計算量表總分: 若有反向題項則反向計分
 - 3.2.2 求出高低分組: 取排序量表總分前後各 25% 處分數
 - 3.2.2 標題: {一些簡單的說明和解釋}
 - ...
- 3.3 (20%) 進行**探索性因素分析** (自行編號及標題，列舉每一做法或步驟，需有一些簡單說明)(參數自選)。綜合上述之分析，哪些題項是需要刪除的? (請寫刪除的理由。)(最後，需印出成份矩陣和因素資料表格)(**採用分析的項目和方法愈多，得分則愈高**)
 - 3.3.1 標題: {一些簡單的說明和解釋}
 - ...
 - 3.3.4 成份矩陣和因素資料表格: 共 X 個因素，我命名為: XXX、XXX 和 XXX。

4 問卷分析: 統計方法 (20%)

資料

同上。

問題

- 4.1 (5%) 不同性別樣本中，對題項 V10(目前的日子過得快樂) 的勾選次數及百分比為何?
- 4.2 (5%) 不同學歷樣本中，對題項 V15(目前的人際關係狀況) 各選項反應百分比是否有顯著不同?
- 4.3 (5%) 不同性別 (只計入男性和女性) 對幸福感 (分析後之分量表總分) 的滿意度是否有顯著不同?
- 4.4 (5%) 不同婚姻狀況對幸福感 (分析後之分量表總分) 是否有顯著不同? 哪些婚姻狀況是顯著不同的?