

108 學年度第二學期
資料採礦: 小考 (1) 第 1 頁/共 3 頁

日期: 2020/04/18(六), 10:30 12:00
授課教師: 吳漢銘 (臺北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. 考試答題要點

- (a) 可參考課本、上課講義 (包含電子檔) 及其它資料。
- (b) 不可與別人 (或同學) 討論，自己做，不可參考同學的答案，不可抄襲。
- (c) 程式設計題，若程式碼直接複製 (或照抄) 講義上的以不給分為原則。
- (d) 請依照「R 程式作業繳交方式」，複製 Console「程式執行及結果」至答案卷。
- (e) 圖形複製，請注意大小，內容數字文字需可辨識。
- (f) 請參照下列文件第 2 ~ 4 頁寫作規定，不按照規定作答者，會扣分。

<http://www.hmwu.idv.tw/web/teaching/doc/R-how-homework.pdf>

2. 下載題目卷，上傳答題檔案:

- (a) 於課程網站下載題目卷。
- (b) 上傳答題檔案: 於教師網站首頁登入 [作業考試上傳區]，帳號: dm108。密碼: xxx (上課教室號碼)。
- (c) 請上傳「學號-姓名-DM-exam1.docx」。(改成自己的學號及姓名)(請注意「正確目錄」)
- (d) 若傳錯，請最終要上傳一份正確的答題檔案。
- (e) 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。
- (f) 若要重覆上傳 (第 2 次以上)，請在檔名最後加「-2」、「-3」，例如: 「學號-姓名-DM-exam1-2.docx」、「學號-姓名-DM-exam1-3.docx」等等。
- (g) 上傳兩次 (含) 以上、格式不合等等酌量扣分。
- (h) 如果上傳網站出現「You can modify the html file, but please keep the link 'www.wftpservers.com' at least.」，請將滑鼠移至「網址列」後，按「Enter」即可。若再不行，請換 (IE/Edge/Firefox/Chrome)。

我已經仔細閱讀上述各注意事項，若有違背，會自行負責。

R: 探索性資料分析及資料處理

資料來源: 政府資料開放平台。資料檔: 「癌症發生統計.csv」。

<https://data.gov.tw/dataset/6399>

1. 請對此資料進行探索性資料分析。
(應至少包含: 列印資料前幾筆紀錄, 各 (連續及類別) 變數之統計圖表、敘述統計, 及對這些結果之簡單說明) (可自行編小題號)
2. 此資料中, 依「縣市別」及「性別 (不分性別、男。女)」之分類, 癌症發生總數各是多少人?
3. 此資料中, 依「性別 (不分性別、男。女)」, 列出癌症發生總數最高之前 5 名的癌症別及其總數。(例: 男性之下, 各癌症別發生總數 = 加總所有「癌症診斷年」及「縣市別」之下的各癌症別發生數)
4. 以「胃癌」為例, 依「性別 (不分性別、男。女)」之類別, 計算每年「年齡中位數」之平均數。