

日期: 2020/10/27(二), 10:30 12:00 上機考: Open Book

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. 考試期間

- (a) 請按照平時上課之座位入座。若同學右邊有人，則該同學應使用隔板 (或拿包包擋著)。
- (b) 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。
- (c) 可使用計算機、自己的筆記型電腦及平板電腦，不可使用手機。
- (d) **全程可上網查詢，但不能用通訊軟體 FB/LINE 等討論。**
- (e) 有問題者，請舉手發問。勿與同學交談。
- (f) 不按照規定作答者，酌量扣分。
- (g) 不可使用它人之隨身碟。「作弊」或「疑似作弊」，往後各項考試不予評分。

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: 於課程網站上登入 [作業考試上傳區]，帳號: fintech。密碼: xxxxx。
- (c) 上傳答題檔案時，請注意「正確目錄」。
- (d) 答題是從「Console」複製程式執行過程及結果，然後貼到答題案。
- (e) 若傳錯，請最終要上傳一份正確的的答題檔案。
- (f) 請上傳「學號-姓名-FinTech-R-exam1.txt」。(學號及姓名，改成自己)
- (g) 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。
- (h) 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-FinTech-R-exam1-2.txt」等等。
- (i) 上傳兩次 (含) 以上、格式不合等等酌量扣分。

3. 完成考試

- (a) 上傳完畢，確認無誤，請刪除作答目錄 及答案卷，清空資源回收筒，並關機。即可離席。

我已經仔細閱讀上述各注意事項，若有違背，會自行負責。

R: 輸入/輸出、程式設計

1. 小吳老師於某系教授 A, B 兩班學生微積分，學期各次成績紀錄於 `Calculus-score-A.csv` 及 `Calculus-score-B.xls` 兩檔案。檔案中紀錄 4 次小考成績、期中期末成績、助教 (TA) 成績，各次考試之配分比例及學期點名出席次數。

- (a) 讀入兩資料檔，將之合併為一個 `data.frame` (命名為 `score`)，使得各欄位名稱如下所示並增加一欄位註明班別 (`Class`)。

```
> score[38:43,]
  Class No   ID Name Gender Quiz1 Quiz2 Quiz3 Quiz4 TA Midterm Final ATT
38    A 38 404550431 沈泓霏   女    15    25    53    67 93.3     29    42    9
39    A 39 404550442 許安霏   女    53    60    80    72 100.0    61    62    9
40    A 40 404550453 李政宜   男    80   100    85   100 100.0    95   100    3
41    B  1 404550465 史文羽   男    60    81   100    97 100.0    90    83    6
42    B  2 404685071 鄭樺妤   男    80   100   100    92 100.0    92    97    2
43    B  3 404685084 張敬安   男    10    40    62    93 100.0    65    84    9
```

- (b) 依各項考試 (小考、期中期末) 配分算出每位同學之學期成績 (缺考以零分計)。其中「出席成績」為額外加分，出席幾次，則總分加幾分。總分以不超過 100 為原則。請列出全班學期成績。
- (c) 列出學期成績在 55~59 分之間的所有同學之全部各欄位紀錄。
- (d) A、B 兩班總成績平均各為多少？男、女生學期成績平均各為多少？
- (e) A 班學期成績不及格比例為多少？B 班男同學學期成績不及格比例為多少？
- (f) 分別印出男、女生學期成績前 5 名之「班別、學號、姓名、學期成績、名次」等欄位紀錄。(男、女生各按照名次依序列出) (名次為全班名次: `rank(x, ties.method = "first")`)
2. 峰度係數 k_c (coefficient of kurtosis) 為一測量峰度高低的量數，可以反映資料的分佈形狀。峰度係數一般是與常態分配作比較而言，該資料分配是否比較高聳或是扁平的形狀。其判別如下：

- 若 $k_c > 0$, 表示資料分布呈高狹峰 (lepto kurtosis)。
- 若 $k_c = 0$, 表示資料分布呈常態峰 (normal kurtosis)。
- 若 $k_c < 0$, 表示資料分布呈低潤峰 (platy kurtosis)。

常用的樣本峰度係數的計算式有以下三項：

- The typical definition used in many older textbooks: $g_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2)^2} - 3$
- Used in SAS and SPSS: $G_2 = \frac{n-1}{(n-2)(n-3)} [(n+1)g_2 + 6]$
- Used in MINITAB and BMDP: $b_2 = (g_2 + 3)(1 - \frac{1}{n})^2 - 3$

其中 n 為樣本大小， x_i 為第 i 個測量值， \bar{x} 為平均數。請寫一 R 函式 (命名為 `my.kurtosis`)，輸入為一組學生成績 (`score`)，輸出為此資料的三項樣本峰度係數。

```
> set.seed(123456)
> score <- rt(150, 4)
> my.kurtosis(score)
$kc
      g2      G2      b2
1.980622 2.089356 1.914436
```

注意: 上傳檔案之後，請刪除作答目錄 及答案卷，清空資源回收筒，關機。