

108 學年度第一學期
電腦概論與程式設計: 作業 (7) 第 1 頁/共 3 頁

繳交日期: 2020/01/06(一), 24:00 前
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. 寫作業要點

- (a) 可參考課本、上課講義 (包含電子檔) 及其它資料。
- (b) 問問題，請多利用課程助教。
- (c) 儘量不要與別人 (或同學) 討論，自己做，不可參考同學的答案，不可抄襲。
- (d) 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。
- (e) 有問題者，請發 e-mail 或 FB 私訊問助教或老師。
- (f) 不按照規定作答者，酌量扣分。
- (g) 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. 上傳答題檔案:

- (a) 於課程網站上登入 [作業考試上傳區]，帳號: r108。密碼: xxxxx。
- (b) 上傳答題檔案時，請注意「正確目錄」。
- (c) 若傳錯，請最終要上傳一份正確的的答題檔案。被移至 {fail} 資料夾的檔案一律不批改。
- (d) 請上傳「學號-姓名-R-HW7.txt」。(學號及姓名，改成自己的)
- (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-R-HW7-3.txt」等等。
- (g) 上傳兩次 (含) 以上、格式不合 (答題檔案內容格式不對、學號多或少一碼、檔名多一空格等等) 酌量扣分。

我已經仔細閱讀上述各注意事項，若有違背，會自行負責。

R: 程式設計

1. 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。

(a) 請寫一函式 (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
```

(b) 讀入資料 score-data.txt 命名為 my.score 物件，使得欄位名稱為科目名，列名稱為學號。利用 apply 及 my.kurtosis 求每一科目的三項樣本峰度係數。

```
> my.score <- ....
> apply(....)
$線代
$線代$kc
      g2      G2      b2
-0.6848024 -0.6282452 -0.7764842
...
...
```

2. 身分證字號驗證規則如下: 字母 (ABCDEFGHIJKLMNPQRSTUVWXYZIO) 對應一組數 (10~35) .

縣市別	英文代號	數字編碼	縣市別	英文代號	數字編碼	縣市別	英文代號	數字編碼
台北市	A	10	新竹縣	J	18	高雄縣	S	26
台中市	B	11	苗栗縣	K	19	屏東縣	T	27
基隆市	C	12	台中縣	L	20	花蓮縣	U	28
台南市	D	13	南投縣	M	21	台東縣	V	29
高雄市	E	14	彰化縣	N	22	澎湖縣	X	30
台北縣	F	15	雲林縣	P	23	陽明山	Y	31
宜蘭縣	G	16	嘉義縣	Q	24	嘉義市	I	34
桃園縣	H	17	台南縣	R	25	新竹市	O	35

令其十位數為 X_1 , 個位數為 X_2 ; (例如 A 的 $X_1 = 1$, $X_2 = 0$) , 令 $D_1 \sim D_9$ 表示第 2~ 第 9 個數字, 再令 $Y = X_1 + 9X_2 + 8D_1 + 7D_2 + 6D_3 + 5D_4 + 4D_5 + 3D_6 + 2D_7 + 1D_8 + D_9$. 如 Y 能被 10 整除, 則表示該身分證號碼為正確, 否則為錯誤。請寫一身分證字號檢查的 R 程式 (命名為 `check.id`) , 輸入為檔名 (`id.txt` 紀錄 5 筆台灣身分證字號) , 輸出為以下表格。

身份字號	數字編碼	縣市別	Y 值	正確性 (Y/N)
=====				
F183741875				
A148992712				
T189179230				
P139392302				
H146359668				