

日期: 2020/11/24(二) 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. 寫作業/考試要點

- (a) 可參考課本、上課講義 (包含電子檔) 及其它資料。
- (b) 考試不可與別人討論，不可抄襲作弊。
- (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) 有問題者，請舉手發問。
- (e) 不按照規定作答者，酌量扣分。
- (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. 上傳答題檔案:

- (a) 於課程網站上登入 [作業考試上傳區]，帳號: r1082。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答題檔案。
- (d) 請上傳「學號-姓名-R-MidtermExam.txt」。(學號及姓名，改成自己)
- (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-MidtermExam-2.txt」等等。
- (g) 上傳兩次 (含) 以上、格式不合、於規定時間之後才上傳等等酌量扣分。

我已經仔細閱讀上述各注意事項，若有違背，會自行負責。

日期: 2020/11/24(二) 上機考: Open Book
授課教師: 吳漢銘 (臺北大學統計學系副教授)

1. 下列 `Letters.code` 為一個包含「A」~「E」的向量。

```
set.seed(123456789)
Letters.code <- sample(LETTERS[1:5], 20, replace=T)
```

- (a) 將 `Letters.code` 中的「A」與「B」編為「第 1 組」,「C」編為「第 2 組」,「D」與「E」編為「第 3 組」。
- (b) 將上小題所得到的編組 `Group.code` 與 `Letters.code` 造成一個資料框 (`data.frame`), 使其具有 `Letters.code` 和 `Group.code` 兩欄位, 且為順序之因子類別。印出此資料之內容及結構。(順序為 A<B<C<D<E; 第 1 組 < 第 2 組 < 第 3 組)
2. (a) 利用 `read_excel {readxl}` 讀取檔案"mydata.xlsx" 的「calculus」工作表。印出資料前後各 5 筆紀錄。
- (b) 將上述讀入資料的欄位名稱重新命名為「No、Department、ID、Name、Quiz1、Quiz2、Quiz3、Quiz4、TA、MidCore1 MidCore2、MidSum」, 並將「ID」指定為列位名稱。(因此資料就沒有「ID」這欄位了。)
- (c) 現各次考試的配分重新指定如下: 4 次小考, Quiz(1)~ Quiz(4), 各佔 8%, 期中考 (MidSum) 佔 30%。期中總成績為上述各次考試之結算, 若缺考以零分計, 請以資料框方式, 列出期中總成績不足 60 分之同學姓名 (Name)、學號 (ID) 及期中總成績, 老師要寄發期中預警單。
3. 資料來源: <https://github.com/owid/covid-19-data/tree/master/public/data>
- (a) 讀取「新冠肺炎」資料 (檔案: `owid-covid-data.csv`), 並印出前後各 5 筆紀錄。
- (b) 計算並印出三個國家 (Germany、United Kingdom 及 United States), 在本年度 8 月份「平均」新增確診案例 (`new_cases`)。
4. 將所觀察到兩變數的資料記做 $\{x_i, y_i\}_{i=1}^n$, 並進行簡單線性迴歸分析。簡單線性迴歸中 ($y = \beta_0 + \beta_1 x + \epsilon$) 之斜率項 (β_1) 及截距項 (β_0) 的估計量如下:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \bar{x} \text{ 和 } \bar{y} \text{ 為 } x_i\text{'s 和 } y_i\text{'s 的平均。}$$

- (a) 試寫一 R 函式, 輸入為兩變數的資料, 輸出為斜率項及截距項的估計量。
- (b) 兩變數之資料如下: `x <- iris[,1]; y <- iris[,2]`, 試用上題之函式計算斜率項及截距項的估計量, 並與 `lm(y~x)` 之結果相比較。

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。交回題目卷。