115-1 程式設計與統計軟體 (期中考) 題目卷

日期: 2026/10/29, 9:20~11:30 (130 分鐘)

注意事項

1. **時間** (請依平時座位入座):
   (a) 考試期間斷網 (9:20~11:30)，僅能使用公用電腦及個人隨身碟作答，不得使用筆電、平板和手機 (關閉網路功能也不行)。
   (b) 考卷下載: 開考 3 分鐘內 (9:20~9:23) 至教學網站下載考卷至公用電腦，將檔案存在自己的隨身碟。
      i. 遲到的同學或斷網後需要 (重新) 下載考卷的同學，請攜隨身碟至教師筆電下載 (會酌量扣分)。
      ii. 下載檔案為: 1151-StatSoft-Midterm.zip，將檔案解壓縮至桌面或隨身碟，產生的資料夾應為「1151-StatSoft-Midterm」。不要跟老師說你不會解壓縮! 不要跟老師說你不知道檔案下載到哪裡去了!
      iii. 裡面包含題目卷、答案卷、資料集 (置於子資料夾「`exam-data`」) 和一些 R 的套件 (zip 檔)，不要跟老師說你不會本地端安裝 R 套件!
      iv. 請用瀏覽器 (Edge/Chrome) 開啟 PDF 檔 (不要跟老師說你不會用瀏覽器開放 PDF 檔!)，不要用 Adobe Acrobat reader (它會一直叫你登入)。
   (c) 考卷上傳: 考後 3 分鐘內 (11:30~11:33) 上傳答案卷。[要提早交卷的同學，請攜隨身碟至教師筆電上傳]
2. 本次考題以 R 程式 (Rgui 或 RStudio) 方式作答，其他程式不允許。
3. 考試過程中可查詢任何資料，包含書本、教學講義、電子檔案 (文件、影音 (請戴耳機))。
4. 禁止利用 FB Messenger, IG, Line 等等通訊軟體與別人討論。禁止使用 chatGPT 或類似之 AI 輔助工具 (Google 搜尋 AI 模式)。禁止 (疑似) 作弊行為。[拜託!! 請不要當害群之馬]。
5. **答案卷**:
   (a) 請使用給定的答案卷樣版做答: MS WORD 檔名為「姓名-StatSoft-Midterm.docx」(「姓名」更改成自己的) 。

(b) 答案卷上請務必於 R Console 內複制「執行後的程式碼及結果 (含圖形)」，於本答案卷貼上 (`Courier New`, 10 點字，白底黑字)，不是只有程式碼，不是只有報表。(老師上課有示範過，請看教學錄影)

(c) 請依序註明題號: 例如: `# ex1(a)`, `# ex1(b)`, `# ex2(a)` 等等。

(d) 若作答不完全，或執行有錯誤，也請於 R Console 內複制「執行後的程式碼及結果 (含圖形)」，貼上答案卷。答題若是空白則無法部份給分。

(e) 題目中的資料檔請以相對路徑讀取 (例如: `"exam-data/Taipei-Population.xlsx"`)，作答前請先將 R 的工作目錄設定為解壓縮後的「1151-StatSoft-Midterm」資料夾 (`setwd` 或 RStudio 專案)。

6. **上傳答案卷**:

(a) 於課程網站上登入 [作業考試上傳區] 或至 `http://hmwu.nccu.edu.tw/login.html?lang=tchinese`

(b) 帳號: statsoft，密碼: xxxx (課程 FB 會公告)，資料夾: 「Exam-Midterm」(考前公告)。

(c) 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，請換其它瀏覽器 (IE/Edge/Firefox/Chrome)。

(d) 上傳檔案，確認檔案大小無誤後，即交卷完畢，無法刪除，也無法再次重傳檔案。若有問題，請連絡老師。

(e) 不要跟老師說你的答題檔案不見了，或變成空白檔了，老師也沒辦法解決。要隨時存檔。

(f) 斷網期間，要提早交卷的同學，請攜隨身碟至教師筆電上傳。上傳檔案後，即可離席。

(g) 考前一天可預先依上述步驟練習上傳任何一個檔案至資料夾: 「上傳測試」。

7. 公用電腦有問題無法解決者，請自行換位至「自由區」或無人使用的「公用電腦」。

8. 有問題者，請先自行想辦法解決，或舉手向老師發問，勿與其它學員連絡討論。

9. 若公用電腦無法讀取個人隨身碟，你告訴老師有問題，老師也沒辦法解決。(平時上課就要試過是否可讀。)

10. 考試現場，老師只有一位，助教只有一位，有問題的同學們請先稍安勿躁，待老師或助教前往了解問題。

11. 考試前有需要的話請先上完廁所。考試期間不得以任何理由隨意進出教室 (除非有特殊狀況)。

12. 請大家一定要先仔細閱讀注意事項，讓考試過程順利圓滿。

R 題目 (6 大題，共 140 分)

說明: 各大題可獨立作答；每一小題皆需呈現「執行後的程式碼及結果」。題目若指定必須使用某函數或指令，請依規定使用；未指定者，可用任何符合題意的 R 寫法。

1. (20 分) 資料檔「exam-data/oath.csv」(UTF-8 編碼，無欄位名稱) 共三列，內容如下:

   ```
   本人
   (系級，學號，姓名)
   恪遵各項考試規則，若如違反，願受校方最嚴厲處罰，謹誓。
   ```

   (a) (10 分) 讀入此資料檔，命名為 oath，印出來 (應為 3 列 1 欄之資料框)。

   (b) (10 分) 利用 readline 印出提示，依序讀入你 (妳) 的系級、學號及姓名，將 oath 第 2 列中的「系級」、「學號」、「姓名」分別置換為所輸入之內容，再將三列文字合併為一個字串，以 cat 印出完整宣誓詞。例如: 依序輸入「統計系二年級」、「1131234567」、「李小華」後，印出:
   本人 (統計系二年級，1131234567，李小華) 恪遵各項考試規則，若如違反，願受校方最嚴厲處罰，謹誓。

2. (10 分) 產生下列向量並印出。(不可用 c() 逐一輸入元素)
   將區間 $[0, 2]$ 等分為 8 個子區間，每一子區間之中點: 0.125 0.375 0.625 $\cdots$ 1.875 (共 8 個)

3. (30 分) Excel 資料檔之讀取與統計運算

   資料檔「exam-data/Taipei-Population.xlsx」(工作表名稱: population) 為臺北市 12 個行政區民國 93 年至 107 年之人口統計資料，每年每區一筆紀錄，共 180 筆。工作表的**前兩列為說明文字，第三列為欄位名稱，第四列起為資料**。欄位依序為: Year(民國年)、Admin.Area(行政區)、Pop.All、Pop.Male、Pop.Female(總人口數、男性及女性人口數，單位: 人)、Household(戶數)、AvePop.H(戶量，人/戶)、Pop.Density(人口密度，人/平方公里)、Gender.Ratio(性比例，每百名女性所對應之男性人數)。

   (a) (10 分) 利用 readxl 套件之 read_excel 讀入此工作表 (**必須跳過前兩列說明文字**，不可手動刪除說明列，也不可先轉存成其它格式)，再以 as.data.frame 轉為資料框，命名為 taipei。印出其維度 (dim)、結構 (str) 及前 3 筆紀錄。

   (b) (10 分) 取出 107 年之資料，依 Pop.All 由大到小排序後，印出 Admin.Area、Pop.All、Pop.Density 三個欄位；並印出 107 年臺北市之總人口數 (12 區 Pop.All 之總和)。

(c) (10 分) 利用 `tapply` 建立一 12×15 之人口密度矩陣 `dens.mat`(列為行政區、欄為年度、元素為 `Pop.Density`；每一 (行政區, 年度) 組合恰有一筆資料，故 `FUN` 可用 `mean` 或 `sum`)，印出其維度及前 3 列前 5 欄，結果格式如下 (行政區列之順序依系統而定，不必相同):

```
> dim(dens.mat)
[1] 12 15
> dens.mat[1:3, 1:5]
          93    94    95    96    97
中山區 15850 15853 16049 15969 15995
中正區 20834 20683 20849 20965 20946
信義區 20745 20591 20653 20421 20323
```

4. (30 分) 迴圈與 `next`、`break`

某社區資源回收 (recycle) 站舉辦「累計回收滿 100 公斤換好禮」活動，自活動第 1 天起連續 20 天之每日回收重量 (公斤) 如下，`NA` 表示該日未營業:

```
12, 8, NA, 15, 9, 11, NA, 14, 10, 13,
7, NA, 16, 12, 9, 10, 11, NA, 8, 14
```

活動規則: 自第 1 天起逐日累計回收重量，未營業之日不計；當累計重量**首次達到 (大於或等於)100 公斤**時，活動於當日結束，之後的資料不再計入。

(a) (5 分) 印出 20 天中有營業的天數，及 20 天之總回收重量 (不含 `NA`)。

(b) (5 分) 不使用迴圈: 將 `NA` 視為 0 後，利用 `cumsum` 印出逐日累計重量之向量 (長度 20)，並利用 `which` 印出累計重量首次達到 100 公斤之天數。

(c) (20 分) 利用 `for` 迴圈 (不可用 `while`/`repeat`，也不可用 `cumsum`) 自第 1 天起逐日處理，且迴圈內**必須使用 `next` 與 `break`**:

- 若當日為 `NA`: 以 `cat` 印出「第 i 天: 未營業, 跳過」，並以一計數變數累計跳過的天數，再以 `next` 跳過該日；
- 否則: 將當日重量累加至變數 `total`，並以 `cat` 印出「第 i 天: 回收 xx 公斤, 累積 xx 公斤」;
- 若 `total` 已達 100 公斤: 印出「第 i 天累積達 100 公斤, 活動結束」，並以 `break` 結束迴圈。

迴圈結束後，印出達標之天數、達標時之累計重量及迴圈結束前實際跳過的天數 (達標天數應與 (b) 之結果一致)。執行結果的前幾列應如下:

```
第 1 天: 回收 12 公斤, 累積 12 公斤
第 2 天: 回收 8 公斤, 累積 20 公斤
第 3 天: 未營業, 跳過
第 4 天: 回收 15 公斤, 累積 35 公斤
...
```

5. (30 分) 自訂函數: 類別變數與連續變數之相關係數 Eta (η)

   設連續變數 y 共 n 筆觀測值 $y_1, \ldots, y_n$，依類別變數 g 分成 k 組 ($k > 2$)，第 j 組有 n_j 筆、平均數為 $\bar{y}_j$，全體平均數為 $\bar{y}$。定義

$$SS_B = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2, \qquad SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2, \qquad \eta^2 = \frac{SS_B}{SS_T}, \qquad \eta = \sqrt{\eta^2},$$

   $0 \le \eta \le 1$，η 越大表示 g 對 y 的解釋力越強。本題使用 R 內建資料集 `iris`(`Species` 為 3 群之類別變數，前 4 欄為連續變數)。

   (a) (5 分) 令 `y <- iris$Sepal.Length`、`g <- iris$Species`。印出各組筆數 n_j、各組平均數 $\bar{y}_j$(利用 `tapply`) 及全體平均數 $\bar{y}$。

   (b) (5 分) η^2 等於線性模型 `lm(y ~ g)` 之判定係數。印出 `summary(lm(y ~ g))$r.squared` 及其平方根。

   (c) (20 分) 撰寫函數 `my_eta(y, g, digits = 4)`: 輸入為連續變數 `y` 與等長之類別變數 `g`，依上述公式計算，回傳一個 `list`，依序包含 `ss.b`、`ss.t`、`eta2`、`eta`(四捨五入至 `digits` 位)；函數內不得使用 `lm`、`aov`、`cor`。印出 `my_eta(y, g)`。

6. (20 分) 二項分配: 程式實作、`d/p/q/r` 與模擬

   小明向心儀的對象告白，假設每次告白成功的機率皆為 $p = 0.25$，且各次告白互相獨立。令 X 為 $n = 10$ 次告白中成功的次數，則 $X \sim B(10,\ 0.25)$，

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

   (a) (10 分) 撰寫函數 `my_dbinom(x, n, p)` 依上述公式計算 $P(X = x)$(`x` 可為向量；不得使用 `dbinom`)。建立並印出一個具有三個欄位的資料框: 欄位 `x`$(0, 1, \ldots, 10)$、`my.pmf`(由 `my_dbinom` 計算)、`dbinom`(R 內建計算)，機率值四捨五入至小數第 4 位。

   (b) (5 分) 利用 `my_dbinom`，計算 10 次告白中至少成功 1 次之機率 $P(X \ge 1)$ 及至多成功 2 次之機率 $P(X \le 2)$。

   (c) (5 分) 利用 `my_dbinom`，找出使 $P(X \le x) \ge 0.9$ 之最小整數 x。