

R程式設計練習題

授課教師: 吳漢銘*
淡江大學數學系資統組

February 10, 2012

Contents

1	範例練習及講解	1
2	平時練習題	3
3	100-1小考 (1)	7
4	100-1小考 (2)	8
5	100-1期末考	9

1 範例練習及講解

每一題, 需要有「輸入參數」的判斷。請注意寫作風格。

1. 輸入包含左右小括號之字串 (最長為 40 字元), 請判斷是否左右小括號配對正確。

(例1) 輸入: $((1+2)-3)*(4/5)$

輸出: 括號配對正確。

(例3) 輸入: $((((1+2+3)$

輸出: 括號配對不正確。

(例3) 輸入: $((1+2)*(3+4)*(5+6))/(7+8)$

輸出: 括號配對正確。

2. 某國發行了1,5,10,50,100不同面額的鈔票, 若有人要從銀行領出 N 元, 銀行行員要如何發給鈔票, 使用的張數會最少?

(例) 輸入: 478

輸出: 1元3張, 5元1張, 10元2張, 50元1張, 100元4張, 共478元。

3. 平面上兩點 (x_1, y_1) , (x_2, y_2) 之的距離式為: $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ 。給定 n 個點($n \leq 10$), 找出構成最小周長的三角形的三個點。

*hmwu@mail.tku.edu.tw

(例) 輸入: (1,1)(0,0)(4,3)(2,0)(7,8)
輸出: 三點為(1,1)(0,0)(2,0), 其周長為4.828428。

4. 輸入任何一個正整數 $n(n \leq 10)$, 輸出 n 階層的 Pascal 三角形。

(例) 輸入: 5

輸出:

```
      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
```

執行程式

```
> source("1234567-Hank-20161220.R")
#####
# Name: 學號-姓名-日期.R          #
#      Exercise                    #
# Author: Han-Ming Wu              #
# Date: 2016/12/20                 #
#####
```

(1): Exercise 1

輸入: 請輸入包含左右小括號之字串 (最長為 40 字元):

...

輸出: xxxxxxxx

繼續練習這一題 (Y)/下一題 (N)? N

(2): Exercise 2

輸入: xxxx

...

輸出: xxxxxxxx

繼續練習這一題 (Y)/下一題 (N)? Y

(2): Exercise 2

輸入: xxxx

...

輸出: xxxxxxxx

繼續練習這一題 (Y)/下一題 (N)? N

(3): Exercise 3

輸入: xxxx

...

輸出: xxxxxxxx

繼續練習這一題 (Y)/下一題 (N)? N

(4): Exercise 4

輸入: xxxx

...

輸出: xxxxxxxx

繼續練習這一題 (Y)/結束 (N)? N

2 平時練習題

1. 丟3顆公平的骰子, 其和為dice.sum,

```
dice.sum <- sum(sample(1:6, 3, replace = TRUE))
```

試寫一函式, 印出總和dice.sum並做如下判別: 如果和大於13點, 則印出「厲害!」, 反之印出「再加油!」。

2. (a) 請利用for寫一函式, 計算一數列之平均數及變異數。
(b) 若有一成績紀錄如下

```
x <- sample(1:100, 50),
```

請利用上小題之函式算出平均數及變異數。

- (c) 請與mean和sd之結果相比較。
3. 利用for寫一函式, 印出九九乘法表。
4. 利用雙迴圈for, 印出下列圖形。

(a)	(b)
1	1
1 2	333
1 2 3	55555
1 2 3 4	7777777
1 2 3 4 5	999999999

5. 利用for, 試計算 $(1 \times 2 \times \dots \times 1000000)$ 之結果所需要的電腦系統時間。
6. 丟一顆公平的骰子 (正六面體)100次, 計算每個數字出現的次數。
7. 大樂透為1~42個號碼一次抽出6個號碼為一組 (取出不放回)。請模擬抽獎100次 (組), 並計算每個數字 (1~42) 出現的次數。
8. (a) 計算 $n!$ 的程式可採用 (1) for, (2) repeat, (3) while, (4) 遞迴法及 (5) R指令 factorial。 (詳細程式見講義)。請用以上五種方法分別計算1000!所需要的系統時間。
(b) 呈上題, 請用指令 system.time 再分別計算一次。
9. 有一50筆成績資料如下

```
score <- sample(1:100, 50, replace = TRUE)
```

判別此資料中是否有高於95分的同學, 若有, 印出「老師請同學吃飯」, 若沒有印出「老師很生氣」。

10. 某班學生期中考微積分及線代的成績資料如下:

```
student.id <- paste("student", 1:50, sep=".")
Calculus <- round(rnorm(length(student.id), mean=65, sd=10), 0)
LinearAlgebra <- sample(1:100, length(student.id), replace = TRUE)
```

- (a) 印出兩科成績皆在85分以上的學生 id。 (Hint: which)
(b) 印出兩科成績皆在60分以下的學生 id。 (Hint: which)
(c) 各科成績最高分及最低分分別是哪些學生? (Hint: max, min)
11. 有一50筆課業成績資料如下

```
score <- sample(1:100, 50, replace = TRUE)
```

大學生課業成績以60分爲及格，以100分爲滿分，而「開根號再乘以10」是著名的成績調分方式，請寫一函式，輸入爲某班學生某科之成績，回傳：(1) 分數調整前被當學生之比例，(2) 分數調整前最高之成績，(3) 分數調整後被當學生之比例，及 (4) 分數調整後最高之成績。

12. 小銘老師有某班學生之期中考試及加分考試兩筆資料，

```
midterm <- sample(1:100, 50, replace = TRUE)
extra <- sample(1:100, 50, replace = TRUE)
```

成績比例爲期中考佔40%，加分考佔60%。結算成績 (100%) 若小於期中考成績，則最後結算成績以期中考計。試寫一函式，處理上述計算，並回傳 (1) 最後結算成績之平均數及變異數，及 (2) 最後被當之學生比例。

13. 某班學生 (student.id) 期中考微積分及線代的成績資料如下：

```
student.id <- paste("student", 1:50, sep=".")
Calculus <- round(rnorm(length(student.id), mean=65, sd=10), 0)
LinearAlgebra <- sample(1:100, length(student.id), replace = TRUE)
```

老師註解成績的方法如下：

- i. 兩科成績皆高於85以上 (含)，記爲「佳」。
- ii. 任一科成績低於40以下 (含)，記爲「要加強」。
- iii. 兩科成績皆低於40以下 (含)，記爲「危險」。

- (a) 利用for寫一函式，計算「佳」「要加強」「危險」各有多少位同學。
- (b) 同一函式裡，再印出「佳」及「危險」之學生座號 (id)。

14. 某班某科原始成績如下:orig.score <- sample(1:100, 55, replace = TRUE)。老師爲了日行一善，打算調整學期總成績 (final.score)，其計算方法有以下三種選擇

- i. 維持原始分數不調分，但高於55分，低於60分者，加至60分及格。
- ii. 「開根號再乘以10」。
- iii. 調成學期總成績最後之平均爲65分，但高於100分者以100計。

試寫一函式，計算 (1) 原始成績之平均數及變異數；(2) 學期總成績之平均數及變異數；(3) 最後被當之學生比例。

15. 某班學生 (student.id) 修課5科成績資料，分別由各科老師提供如下：

```
student.id <- paste("student", 1:55, sep=".")
set.seed(123)
Calculus <- round(rnorm(length(student.id), mean=65, sd=10), 0)
LinearAlgebra <- sample(1:100, length(student.id), replace = TRUE)
BasicMath <- sample(1:100, length(student.id), replace = TRUE)
Rprogramming <- sample(1:100, length(student.id), replace = TRUE)
English <- sample(1:100, length(student.id), replace = TRUE)
```

- (a) 請將此各別資料轉成單一資料表格 (命名爲mydata)，使得欄位名稱爲科目名，列名稱爲學生的座號student.id，並列印出前3位同學成績紀錄。
- (b) 請將資料依LinearAlgebra排序後，印出此科目最高分及最低分各5位同學的各科成績。
- (c) 若每科學分數皆爲3學分，同時每科以60分爲及格。請找出1/2的同學。

16. 某班學生 (student.id) 某科期中考成績 (score) 資料如下:

```
student.id <- paste("student", 1:50, sep=".")
my.p <- dnorm(seq(-3,3,length=100))
set.seed(123456)
score <-sample(1:100, length(student.id), replace = TRUE, prob=my.p)
```

大學生課業成績以60分爲及格, 以100分爲滿分, 請寫一函式, 以「開根號再乘以10」爲調分方式, 輸入爲某科之成績, 回傳:

- (a) 分數調整前, 不及格學生之比例。
- (b) 分數調整前, 最高成績之學生座號。
- (c) 分數調整後, 全班成績之平均數及標準差。

17. 公式

$$z = \frac{x - \bar{x}}{s} \quad (\text{其中 } \bar{x} \text{ 爲平均數, } s \text{ 爲標準差})$$

一般稱做 z -轉換 (z -transformation) 或 z -分數 (z -score)。

- (a) 請寫一函式, 輸入爲一數列, 輸出爲 z 分數。
- (b) 若有5個成績 $x \leftarrow \text{sample}(1:100, 5)$, 請利用上述之函式轉成 z 分數, 並算出轉換後之平均數及變異數。
- (c) 請與scale之結果相比較。

18. 相關係數之公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- (a) 試寫一函式, 計算兩變數之相關係數。
- (b) 兩變數之資料如下: $x \leftarrow \text{rnorm}(20)$; $y \leftarrow \text{runif}(20)$, 試用上題之函式計算相關係數, 並與cor之結果相比較。

19. 截尾平均數 (trimmed mean) 是將一資料排序後, 將頭尾拿掉一定百分比 p 的觀察值, 然後用剩下的觀察值計算平均數。

- (a) 試寫一函式, 計算截尾平均數。
- (b) 若有兩資料data1和data2如下:

```
data1 <- rnorm(100)
data2 <- data1[sample(100, 10)] + 2*qchisq(0.975, 10)
```

呈上題, 計算兩資料之截尾百分比 p 爲1%、3%、5%及10%之截尾平均數。

- (c) 請與 $\text{mean}(\text{data1}, \text{trim} = p)$, $\text{mean}(\text{data2}, \text{trim} = p)$, 其中 $p=0.01, 0.03, 0.05, 0.1$ 的結果相比較。

20. 牛頓法求 $f(x) = 0$ 的解 (Newton's Method) 過程如下:

先猜測一初始值 x_0 爲 $f(x) = 0$ 的根, 再以初始值 x_0 代入下列迭代法求得第一次近似根 x_1 , 如此一直重覆此過程而得到近似解:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad \text{if } f'(x_n) \neq 0.$$

- (a) 試寫一函式, 利用牛頓法求 $f(x) = 0$ 之解。
- (b) 呈上小題, 求解 $f(x) = x^2 - 2 = 0$ 的正根, 初始值爲 $x_0 = 1$ 。(至少迭代5次以上)
- (c) 請與 $\text{sqrt}(2)$ 之結果相比較。

21. 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:

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

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

- (a) 試寫一函式, 計算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。

3 100-1小考 (1)

- 讀入資料`score-data.txt`(其類別為 `data frame`), 命名為`my.score`物件, 使得欄位名稱為科目名, 列名稱為學號。列印出前5位同學所有成績紀錄。
 - 將資料`my.score`的列 (`row`) 命名為 `student.1`, `student.2`, \dots , `student.n` (n 為 `row` 的個數)。
 - 將「基數」的成績以「開根號乘以十」重新計算後, 結合全班其它各科成績匯出成另一資料檔`new-score.txt`, 內容需有欄位名稱, 列名稱, 並以 `TAB` 作分隔, 而且輸出資料不要有引號。
- 寫一函式 (`my.test`), 輸入為一組學生成績 (`score`), 判別此資料, 若「成績及格人數達半數以上 (含), 且有90分以上 (含) 之同學」則印出「本次成績不調分, 平均為: `xx.xx`」否則印出「本次成績會調分, 不及格比例為: `xx.xx`」。(小數點以下兩位)

```
> set.seed(123456)
> score <- sample(1:100, 50, T)
> my.test(score)
本次成績不調分, 平均為: 55.78
>
> set.seed(123456)
> score <- sample(1:100, 150, T)
> my.test(score)
本次成績會調分, 不及格比例為: 60.67 %
```

- 寫一函式, 使其執行時會要求使用者輸入一 3×3 矩陣, 在計算其反矩陣後, 回傳原始矩陣及其反矩陣。(提示: `solve`)。請利用以下矩陣做測試:

$$\begin{bmatrix} 3 & 5 & -1 \\ 2 & -1 & 3 \\ 4 & 2 & 3 \end{bmatrix}$$

4 100-1小考 (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。

1. 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
```

2. 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
...
...
```

5 100-1 期末考

1. 簡單線性迴歸模式中 ($y = \beta_0 + \beta_1 x + \epsilon$) 之斜率項 (β_1) 及截距項 (β_0) 的估計量如下:

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

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

若兩變數之資料如下: `x <- iris[,1]`; `y <- iris[,2]`, 試寫一R函式`my.lm(x, y)`, 計算斜率項及截距項的估計量。並與`lm(y~x)`之結果相比較。

2. 樂透彩對獎程式: 在1~42 的整數中, 樂透彩會開出 6 個號碼以及一個特別號, 中獎規則以及獎額如下:

獎項	規則	獎金
頭獎	6 個號碼全中	1,000,000
二獎	6 個號碼中 5 個, 另一個中 特別號	100,000
三獎	6 個號碼中 5 個	10,000
四獎	6 個號碼中 4 個	1,000
五獎	6 個號碼中 3 個	100

註: 中頭獎的不能再被視爲中三獎, 餘類推。

- (a) 若當期開出之號碼爲 38, 28, 18, 8, 5, 10。而特別號是 42。小銘買了一張彩卷, 選號爲 15, 1, 8, 18, 28, 38。請問有對中之號碼爲何? 對中號碼個數爲幾個?
- (b) 小吳也買了樂透彩, 所選 5 組號碼記錄在 (`mylist.txt`) 檔案。請你寫一R程式`lotto`幫他對獎, 使得輸出爲以下所列。(提示 1: 輸入爲當期開出之號碼、特別號及號碼記錄檔。)

no.	中獎	累積獎金	有對到之號碼
1	沒中獎	0	(38)
2	中二獎	100000	(8 18 28 38 5 [42])
3	中五獎	100100	(8 18 28)
4	沒中獎	100100	()
5	中頭獎	1100100	(5 10 8 18 28 38)
=====			
總獎金: 1100100 元			

(提示 2: `as.matrix`, `as.integer`, `if`, `for`, `cat`, `length`, `which`, ...)