

# 中央研究院 AI 跨域領袖營: R 語言專題實作練習

2017/09/22(五)

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

### 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) 可與別人討論，但不可抄襲。
- (c) 有問題者，請發 e-mail 或 FB 私訊問老師。
- (d) 請參照教學網站上文件「R 程式作答方式」寫作規定。
- (e) 記得寫題號。
- (f) 貼圖是圖形本身，不是螢幕截圖畫面，也不是軟體截圖畫面。

### 2. 上傳答題檔案:

- (a) 於教師網站首頁 (<http://www.hmwu.idv.tw/>) 登入 [作業考試上傳區]，  
帳號: ai-r。密碼: xxxx (上課日期)。
- (b) 請上傳「姓名-R-exercise.docx」。(姓名，改成自己)
- (c) 若要重覆上傳 (第 2 次以上)，請在檔名最後加「-2」、「-3」，例如: 「姓名-R-exercise-2.docx」、「姓名-R-exercise-3.docx」等等。

# 中央研究院 AI 跨域領袖營: R 語言專題實作練習

2017/09/22(五)

授課教師：吳漢銘 (臺北大學統計學系副教授)

## 1 基礎題

1.1 假設作業系統是 MS Windows，並使用 Rgui 或 RStudio。

- (a) 印出系統現在之年月日及時間。
- (b) 列出目前的工作目錄。
- (c) 列出目前目錄下的子目錄及檔案 (提示: `list.dirs`, `list.files`)。
- (d) 重新設定工作目錄在「C:\Users\Default」。

1.2 安裝套件 (`install.packages`)

- (a) 從台大資工 CRAN 鏡射站安裝兩個套件"cluster, clValid"，並載入 R。
- (b) 到此位置<https://cran.r-project.org/web/packages/seriation/index.html> 下載 seriation 套件至電腦中。並在 Rgui 或 RStudio 中以 `install.packages` 指令安裝。
- (c) 在 Rgui 或 RStudio 中安裝三個 Bioconductor (<https://bioconductor.org>) 套件: `cancerclass`, `geneClassifiers`, `maSigPro`。

1.3 列出電腦作業系統 (含位元數) 及 R 版本等等系統資訊。

1.4 某學生分析空氣品質資料 `airquality` 之風速 (Wind) 與溫度 (Temp) 的關係，他採用迴歸分析及變共數分析，步驟如下：

```
lm.obj <- lm(airquality$Wind ~ airquality$Temp)
lm.anova <- anova(lm.obj)
lm.summary <- summary(lm.obj)
```

- (a) 物件 `lm.anova` 是屬於何種類別，其儲存結構如何？
- (b) 物件 `lm.summary` 有哪一些屬性可供存取？試取出  $R^2$  值。(提示: `r.squared`)

1.5 由螢幕輸入 2 個數字 (例如: 26, 87)，印出其總和。

1.6 某班學生有一考試成績和性別紀錄如下 (資料是依照學生座號 1、2、... 依序紀錄; NA 代表缺考):

```
成績: 30, 49, 95, NA, 54, NA, 61, 85, 51, 22, 0, 0
性別: m, f, f, m, f, m, f, m, m, f, f, m
```

- (a) 本班共有多少學生？男女生各多少人？
- (b) 此科目成績最高分及最低分是幾分？

- (c) 計算此科目成績平均及標準差。男生成績平均各是多少？
- (d) 老師欲將成績依序做以下調整: (i) 缺考以 0 分計;(ii) 每人加 10 分 (缺考者不加分，超過 100 分以 100 分計)。印出調整後的分數。
- (e) 以調整後的分數計，列出及格 (60 分以上，含) 同學的座號。共有幾位？

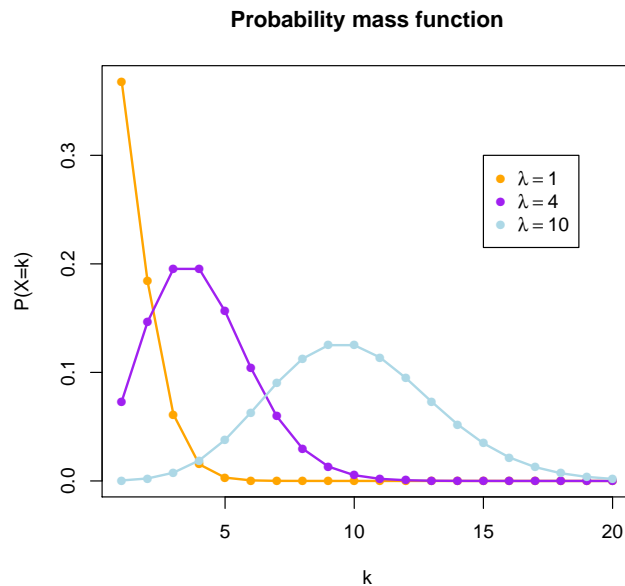
1.7 某班某次考試之成績 (ScoreData) 如下，(a) 試計算每人之平均分數。(b) 若三科成績 (math, english, algebra) 計算平均之權重依序為 (0.5, 0.2, 0.3)，試計算每人之加權平均分數。(提示: apply, mean, weighted.mean)

```
set.seed(123456789)
math <- sample(1:100, 10, replace=T); english <- sample(1:100, 10, replace=T)
algebra <- sample(1:100, 10, replace=T); ScoreData <- cbind(math, english,
algebra)
```

1.8 卜瓦松分布 (Poisson distribution) 的機率質量函數 (Probability mass function) 為

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

利用 matplotlib 畫出機率質量函數圖如下 ([https://en.wikipedia.org/wiki/Poisson\\_distribution](https://en.wikipedia.org/wiki/Poisson_distribution))。(提示: c("orange", "purple", "lightblue"), data.frame, type="o", expression)



## 2 挑戰題

2.1 某班「R 程式設計」一科學期各項成績總表紀錄於「R-score.xlsx」。

- (a) 讀取資料檔，印出前 5 位同學成績紀錄。
- (b) 計算各項考試 (不含點名) 平均分數及標準差。
- (c) 依照各項考試配分 (小考 1(10%)，小考 2(15%)，小考 3(15%)，作業 (20%)，期末考 (40%)) 計算每位同學之學期成績，並以 `data.frame` 的類別型式印出學號及學期成績。(其它項目不用列出)

2.2 某樂透 (Lottery) 遊戲規則如下：「消費者從 01~49 中任選 6 個號碼進行投注。開獎時，開獎單位將隨機開出 6 個號碼 (winning number)。如果消費者選號有三個以上 (含三個號碼) 對中當期開出之 6 個號碼，即為中獎，並可依規定兌領獎金。」某天小明買了兩注電腦選號，其號碼為 (5, 29, 12, 10, 38, 35) 和 (41, 13, 21, 29, 19, 12)。若當期之開獎號碼為 (10, 7, 12, 38, 47, 35)，請寫一 R 函式，幫小明對獎。程式要求如下：(1) 輸入為開獎號碼 (預設值為本題之開獎號碼)；(2) 執行對獎程式後，由螢幕輸入「消費者投注號碼」；(3) 輸出為消費者投注號碼及開獎號碼、對中之號碼個數、恭喜中獎或銘謝惠顧；(3) 不可用 `for`。(提示：(1) `%*%`；(2) 由螢幕輸入「消費者投注號碼」，可一次輸入兩注，或一次輸入一注但執行兩次對獎程式)

2.3 美國大學成績平均績點 (GPA)(四分制) 的計算方式如下表：

等級 (Grade)	百分數	GPA
A	80 – 100 分	4
B	70 – 79 分	3
C	60 – 69 分	2
D	50 – 59 分	1
E	49 分以下	0

請寫一 R 函式，將某同學之各科修課成績百分數 (`score`) 轉成等級及 GPA。(提示：不可用 `for`)

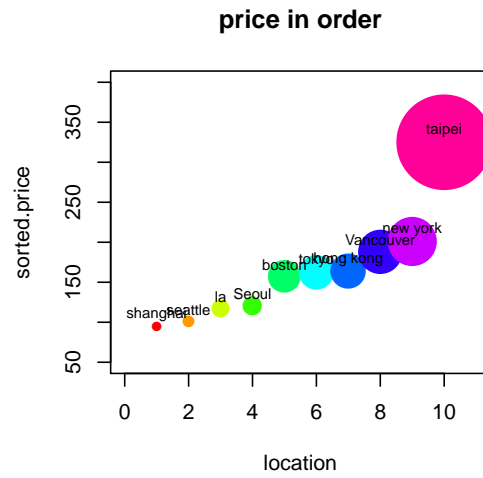
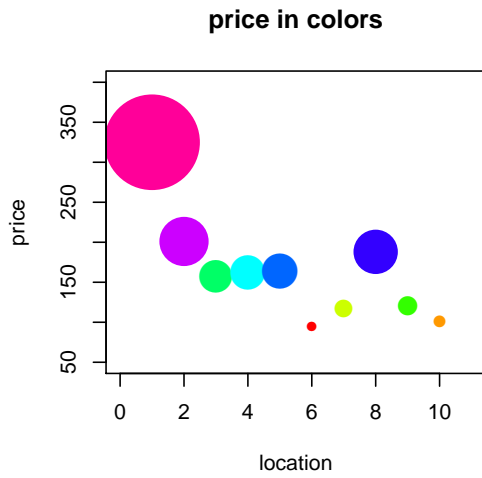
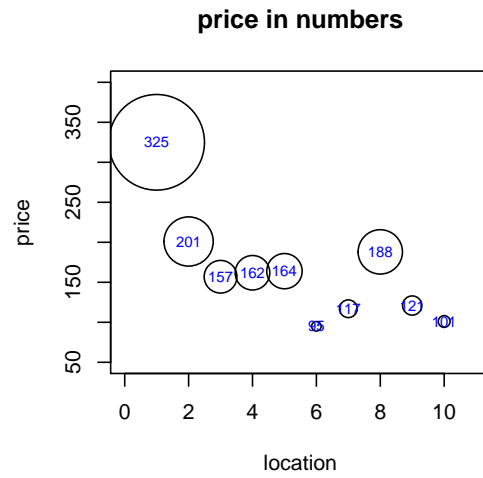
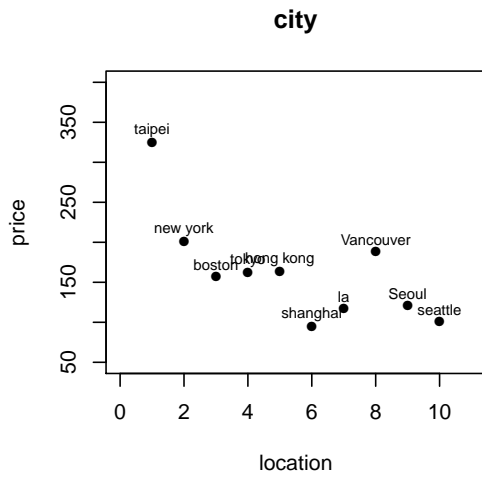
```
> set.seed(12345)
> score <- sample(0:100, 10, replace=T)
```

2.4 讀取資料檔 `city.txt`：

- (a) 依照下列公式，將 `price` 之值轉換為範圍介於 (1~10) 之值，並列印出來。

$$\text{transformed.price} = 9 \times \frac{\text{price} - \min(\text{price})}{\max(\text{price}) - \min(\text{price})} + 1$$

- (b) 繪出下列一頁 4 圖 (其中 Bubble plot 之泡泡大小是依據上小題)。(提示：`sort`, `order`, `rainbow(10)`)



2.5 ChickWeight {datasets} 是 R 內建的資料。紀錄小雞在不同飼料餵食之下的體重。關於此資料更多說明，請「?ChickWeight」。利用所學之統計圖探索此資料。(不需解說所畫出之圖形所代表的意義。)(圖之標題， $xy$  軸標號及圖例說明需完整，讓讀者一看就可得知資訊; 一張圖以呈現最多資訊為原則)