**實例練習1**

將sex及age共10筆資料輸入R中，並簡單分析資料內容。

【性別：女生為0，男生為1】

R統計軟體輸出或輸入的格式為文字檔(.txt)或是格式檔(.csv)的型態

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sex** | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| **age** | 56 | 45 | 62 | 68 | 45 | 40 | 38 | 51 | 65 | 45 |

# 設定工作目錄

setwd("C:/RData")

資料一律放在此目錄下

步驟一

方法一：利用語法輸入資料

|  |  |
| --- | --- |
| R語法 | 語法解釋 |
| (直接將資料輸入於R中)>a<-matrix(0,10,2)>a[1:10,1]<-c(0,0,0,0,1,1,1,1,1,1)>a[1:10,2]<-c(56,45,62,68,45,40,38,51,65,45)>rownames(a)=c(1:10)>colnames(a)=c(“sex”,”age”)>a | 註：在”>”之後輸入語法# 建立一個10x2的0矩陣a，matrix(資料,列,行)# 宣告第一行的內容# 宣告第二行的內容( c(數字,數字,….. )， c(“文字”,”文字”,….) )# 矩陣列的名成設為1~10# 矩陣行的名稱設為”sex”,”age”# 查看矩陣a的內容 |

* **結果：**



方法二：讀取外部檔案，利用文字檔或格式檔建立資料

|  |  |
| --- | --- |
| R語法 | 語法解釋 |
| (以文字檔型態輸入)>hello<-read.table("c:/RData/hello.txt", header=T) | # 讀取儲存的文字檔(\*.txt) |
| (以格式檔型態輸入)>hello<-read.csv("c:/RData/hello.csv") | # 讀取儲存的格式檔(\*.csv) |

步驟二：利用tapply列出描述性統計值：

|  |  |
| --- | --- |
| R語法 | 語法解釋 |
| >attach(hello)>tapply(age,sex,summary)>tapply(age,sex,length) | # 連結資料hello# tapply:以類別變數分組，列出連續變數之描述性統計值，語法如下:tapply(連續變數名, 類別變數名, 指定功能參數)指定功能參數:summary：要求列出基本描述性統計值 length: 要求列出每組個數ps.若是延續方法一的步驟，需將資料存成.csv或.txt才可以進行資料連結# 依據sex分類算出個別的平均值、個別長度等變項，tapply(連續變項,類別變項,函數) ，tapply的變項亦可為矩陣 |

* 結果：

> tapply(age,sex,summary)

$`0`

 Min. 1st Qu. Median Mean 3rd Qu. Max.

 45.00 53.25 59.00 57.75 63.50 68.00

$`1`

 Min. 1st Qu. Median Mean 3rd Qu. Max.

 38.00 41.25 45.00 47.33 49.50 65.00

> tapply(age,sex,length)

0 1

4 6

資料整理如下：

|  |  |  |
| --- | --- | --- |
|  | 男生(1) | 女生(0) |
| 個數 | 6 | 4 |
| 平均值 | 47.33 | 57.75 |
| 最大值 | 65.00 | 68.00 |
| 最小值 | 38.00 | 45.00 |

利用tapply計算出R資料的簡單描述性統計值。從表格中可知，女生共有4人，其平均年齡約在57.75歲；男生共有6人，其平均年齡約在47.33歲，亦可由表格中得知最大值及最小值。