



112-2 生物統計學一

# 資料建立

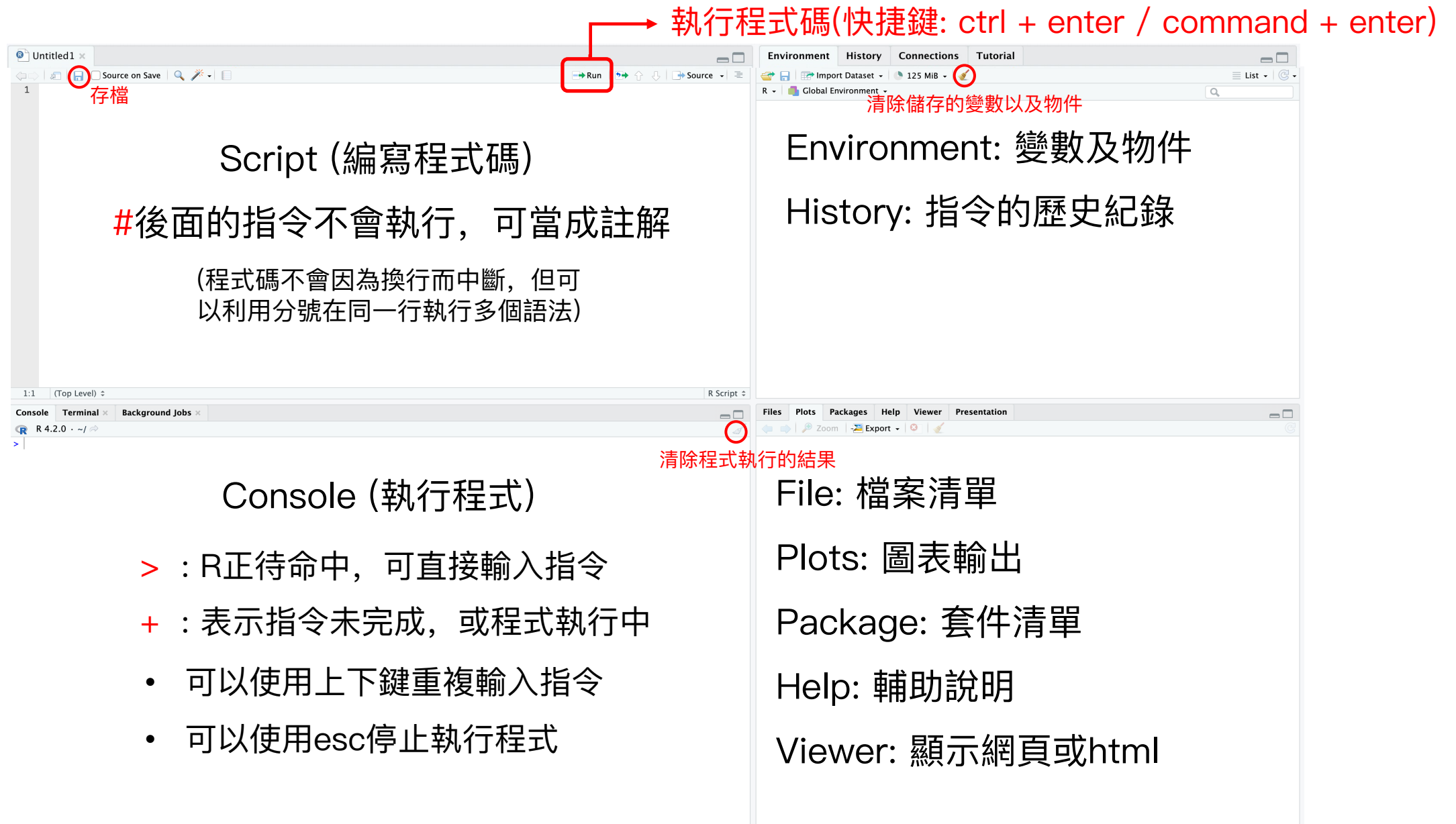
2024/02/27

助教: 廖振博

## PART 01

# 基本功能介紹

# RStudio的介面



The screenshot shows the RStudio interface with several components and annotations:

- Script (編寫程式碼):** The main editor window. Annotations include:
  - 存檔 (Save):** A red circle around the save icon in the top-left toolbar.
  - 執行程式碼 (快捷鍵: ctrl + enter / command + enter):** A red circle around the "Run" button in the top-right toolbar, with an arrow pointing to the text.
- Environment:** The top-right pane, showing variables and objects. Annotation: **清除儲存的變數以及物件 (Clear stored variables and objects)** with a red circle around the trash icon.
- Console (執行程式):** The bottom-left pane. Annotation: **清除程式執行的結果 (Clear execution results)** with a red circle around the trash icon.
- Other panes:** The bottom-right pane contains a list of other RStudio features: File: 檔案清單 (File: File list), Plots: 圖表輸出 (Plots: Plot output), Package: 套件清單 (Package: Package list), Help: 輔助說明 (Help: Help), and Viewer: 顯示網頁或html (Viewer: Display web page or html).

# R的物件命名規則

## ❶ 為什麼需要命名，什麼是物件？

- R是一種物件導向的程式語言，通常以物件的操作來執行程式
- 物件可以是數字、資料、函數、模型、分析結果.....

## ❷ 物件命名規則：

- 只能以字母、數字、底線、小數點組成物件名稱
- 開頭必須是英文字母或小數點
- 以小數點開頭的物件，小數點後只能是字母
- 大小寫不同(跟SAS不同)
- 不可以使用R中的保留字命名，例如if與NA

## ❸ 指定物件名稱或賦值的方式：

```
> x <- 5  
> x = 5  
> 5 -> x  
> assign("x", 5)
```

# R常見的物件/資料型態

- ① 字串(character string)/數值(numeric):

```
> (aa <- "AA")  
[1] "AA"  
  
> (bb <- 100)  
[1] 100
```

- ② 向量(Vector): 用來儲存**單一維度**的資料，只能包含一種類型的資料(例如數值或字串)

```
> (cc <- c(1, 3, 4, 5, 6, 7))  
[1] 1 3 4 5 6 7
```

- ③ 矩陣(Matrix): 用來儲存**兩個維度**的資料，只能包含一種類型的資料

```
> (dd <- matrix(cc, 2, 3, byrow = T))  
      [,1] [,2] [,3]  
[1,]    1    3    4  
[2,]    5    6    7
```

- ④ 資料框(Data frame): 用來儲存**兩個維度**的資料，可以包含**多種**類型的資料

```
> (ee <- data.frame(Name = c("John", "Alice"), Age = c(25, 30)))  
  Name Age  
1 John  25  
2 Alice 30
```

- ⑤ 列表(Data frame): 用來儲存**很多個**物件，可以包含向量、矩陣、資料框、甚至其他列表

```
> (ff <- list(c(1, 2, 3), c("A", "B")))  
[[1]]  
[1] 1 2 3  
  
[[2]]  
[1] "A" "B"
```

# 物件操作的四個通則(CRUD)

---

- ❶ C (create): 建立物件
- ❷ R (read): 讀取物件裡的元素
- ❸ U (update): 更新/修改物件裡的元素
- ❹ D (delete): 刪除物件或刪除物件裡的元素

# 向量(Vector)

## ① C (create): 建立物件

```
> (xx <- c(1, 2, 3, 4, 5))  
[1] 1 2 3 4 5
```

## ② R (read): 讀取物件裡的元素

```
> xx[2]  
[1] 2
```

## ③ U (update): 更新/修改物件裡的元素

```
> xx[2] <- 100
```

## ④ D (delete): 刪除物件或刪除物件裡的元素

```
> xx[-3]  
[1] 1 100 4 5
```

# 矩陣(Matrix)

## 1 C (create): 建立物件

```
> (yy <- matrix(data = c(1:6), nrow = 2, ncol = 3, byrow = FALSE))
```

	[,1]	[,2]	[,3]
[1,]	1	3	5
[2,]	2	4	6

## 2 R (read): 讀取物件裡的元素

```
> yy[1, 2]
```

[1]	3
-----	---

## 3 U (update): 更新/修改物件裡的元素

```
> yy[1, 3] = 100
```

```
> yy
```

	[,1]	[,2]	[,3]
[1,]	1	3	100
[2,]	2	4	6

## 4 D (delete): 刪除物件或刪除物件裡的元素

```
> yy[-1, -3]
```

[1]	2	4
-----	---	---



# 練習1

- 建立矩陣  $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$
- 將上方的矩陣變成  $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 1000 & 9 \end{pmatrix}$
- 利用上方的矩陣生成  $\begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}$

# 資料框(Data Frame)

## 1 C (create): 建立物件

```
> id <- c(1:4)
> age <- c(25, 30, 35, 40)
> sex <- c("male", "male", "female", "female")
```

```
> (zz = data.frame(id, age, sex))
  id age  sex
1  1  25 male
2  2  30 male
3  3  35 female
4  4  40 female
```

## 2 R (read): 讀取物件裡的元素

```
> zz$sex
[1] "male" "male" "female" "female"
> zz[, 3]
[1] "male" "male" "female" "female"
```

## 3 U (update): 更新/修改物件裡的元素

```
> zz$age <- c(10, 10, 10, 10)
```

## 4 D (delete): 刪除物件或刪除物件裡的元素

```
> zz[, -1]
```

# 列表(List)

- 1 C (create): 建立物件

```
> (ww = list(xx, yy, zz))
```

- 2 R (read): 讀取物件裡的元素

```
> ww[[3]]
```

```
> ww[[3]]$age
```

- 3 U (update): 更新/修改物件裡的元素

```
> ww[[3]] = c(1,2,3)
```

- 4 D (delete): 刪除物件或刪除物件裡的元素

```
> ww[[2]] <- NULL
```

## PART 02

# 資料匯入與匯出

# 資料匯入(txt檔、csv檔)

## ① txt檔(常為空白分隔)

```
> read.table(file = , header = TRUE, sep = "")
```

- file(檔案路徑): 檔案路徑(注意斜線方向), 路徑需加上引號或使用file.choose()點選
- header(欄位標題): 第一列是否為標題
- sep(分隔符號): 檔案由什麼符號分隔

## ② csv檔(逗點分隔)

```
> read.csv(file = , header = TRUE, sep = ",")
```

# 查看資料內容

- 1 查看整個資料

```
> View()
```

- 2 顯示前/後幾筆資料，預設為6筆

```
> head() ; tail()
```

- 3 查看資料欄列數

```
> dim()
```

- 4 資料欄或列的名稱

```
> colnames() ; rownames()
```

# 資料匯出(txt檔、csv檔)

## ❶ txt檔(常為空白分隔)

```
> write.table(x, file = , row.names = TRUE, sep = " ")
```

- x(資料表格): R內部資料的名稱
- file(檔案路徑): 匯出檔案路徑
- row.names(列標籤): 每列名稱是否要匯出
- sep(分隔符號): 檔案是由什麼符號分隔

## ❷ csv檔(逗點分隔)

```
> write.csv(x, file = , row.names = TRUE)
```

# 練習2

---

- 將Studata.csv匯入R中，並指定名稱為indata
- 保留編號(ID)、年紀(Age)、性別(Gender)、身高(Ht)、體重(Wt)，並命名為outdata
- 將outdata匯出至外部，並指定匯出的資料為practice1.csv