



112-2 生物統計學一

資料整理、篩選

2024/03/12

助教: 廖振博

PART 01

複習

物件操作的四個通則(CRUD)

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]
```

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 = ",")
```

資料匯出(txt檔、csv檔)

① 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

② csv檔(逗點分隔)

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

設定工作目錄(work directory)

> `getwd()`

- 顯示當前的工作目錄

> `setwd("目標路徑")`

- 設定當前的工作目錄
- 需先確認目標路徑存在，不會創建不存在的目錄
- 關掉R之後又會回到原來預設的路徑

PART 02

數學運算

R的數學運算

- 加減乘除

```
> (8 + 9 - 10) * 3 / 10
[1] 2.1
```
- 次方

```
> 4 ^ 3 # 4 ** 3 也可以
[1] 64
```
- 開根號

```
> sqrt(81)
[1] 9
```
- 商數

```
> 11 %/% 5
[1] 2
```
- 餘數

```
> 11 %% 5
[1] 1
```
- 還有exp(), log(), abs(), sign(), round().....

常見的幾個計算函數

```
> xx <- c(1:100)
> sum(xx)
[1] 5050
> var(xx)
[1] 841.6667
> mean(xx)
[1] 50.5
> max(xx)
[1] 100
> min(xx)
[1] 1
```

```
> (yy <- matrix(c(1:25), 5, 5, byrow = TRUE))
      [,1] [,2] [,3] [,4] [,5]
[1,]     1     2     3     4     5
[2,]     6     7     8     9    10
[3,]    11    12    13    14    15
[4,]    16    17    18    19    20
[5,]    21    22    23    24    25
> rowSums(yy)
[1] 15 40 65 90 115
> rowMeans(yy)
[1] 3 8 13 18 23
> colSums(yy)
[1] 55 60 65 70 75
> colMeans(yy)
[1] 11 12 13 14 15
```

練習1

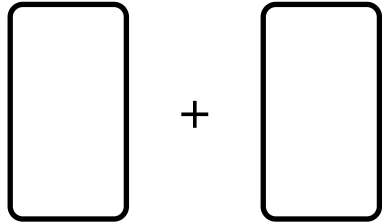
- 將Studata.csv匯入R中，並指定名稱為indata
- 計算每個人的BMI，並命名為bmi新增到outdata中。
- 計算每位同學這學期(credit_current)與上學期(credit_last)修課學分數的差值，並命名為credit_difference新增到outdata中。
- 將outdata匯出至外部，並指定匯出的資料為practice1.csv

PART 02

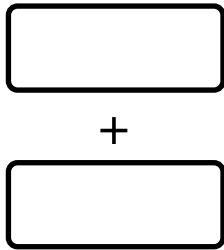
資料合併、篩選

資料合併

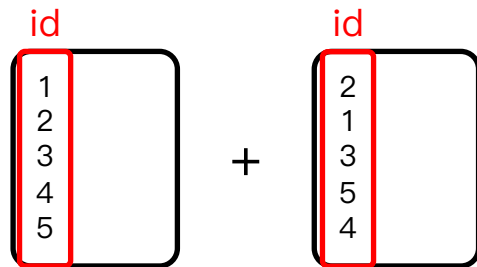
- `cbind(資料1,資料2,)`: 行資料合併



- `rbind(資料1,資料2,)`: 列資料合併



- `merge(資料1, 資料2, by = “共同欄位”)`: 欄資料合併(根據共同欄位合併)



邏輯判斷

- 常見的邏輯判斷符號：
 - $>$ 、 $<$: 大於、小於
 - \geq 、 \leq : 大於等於、小於等於
 - $==$ 、 $!=$: 等於、不等於
 - $A \%in\% B$: A 是否在 B 中
 - $\&$: 交集(and)
 - $|$: 聯集(or)
- ifelse(條件, 滿足條件時的值, 不滿足條件時的值)

- which(邏輯判斷)

```
> which(data1$BMI > 40)
```

```
[1] 6 8 11 35 42 63 84 103 124 134 144 187 193 231 232  
[16] 251 257 292 296 312 321 324 334 343 346 376 384 389 402 420  
[31] 480 508 517 519 548 550 601 629 640 644 652 695 742 752 767  
[46] 771 781 826 836 837 838 863 868 884 907 949 987 1048 1054 1068  
[61] 1082 1083 1096 1115 1162 1173 1230 1246 1256 1258 1282 1287 1382 1383 1387  
[76] 1463 1475 1518 1529 1573 1582 1590 1594 1620 1642 1689 1721 1727 1914 2151  
[91] 2253 2305 2330 2331 2338 2491 2493 2511 2646
```

- subset(資料, 邏輯判斷, select = 欲選取欄位)

```
> subset(data1, BMI > 40, select = c(age, diabetes, BMI))
```

	age	diabetes	BMI
6	46	yes	46.26
8	46	no	40.62
11	47	no	44.75
35	51	yes	44.38
42	52	no	40.06
63	53	no	45.04

練習2

- 利用練習1計算的bmi依下列方式進行分組，並指定新變數名稱為bmi_group
 - 0: $\text{BMI} < 18.5$
 - 1: $18.5 \leq \text{BMI} < 24$
 - 2: $\text{BMI} \geq 24$
- 將bmi_group作為新變數，新增到indata中
- 計算全班、男生($\text{Gender} = 1$)與女生($\text{Gender} = 0$)身高(Ht)的平均值

本日作業

- 生統實習課__作業1 (作業區)
- 所需資料檔： hwdata.csv
- 繳交期限： 2024.03.19 (二) 23:59 前上傳至COOL作業區
- 遲交一天扣10分，請務必確認是否上傳成功
- 作業格式： pdf 檔