

R 基本實作

秘訣速查表

查詢使用說明

R環境的查詢指令

`?mean`
取得函式的說明文件

`help.search('weightedmean')`
以關鍵詞查詢說明文件

`help(package = 'dplyr')`
查詢套件的說明

查詢已存物件的資訊

`str(iris)`
顯示物件的資料結構

`class(iris)`
顯示物件的資料形態

使用套件

`install.packages('dplyr')`
從CRAN伺服器下載並安裝套件

`library(dplyr)`
載入已安裝套件到工作環境·直接使用套件內函式

`dplyr::select`
不事先載入套件·直接使用套件內函式

`data(iris)`

工作目錄

`getwd()`
顯示目前工作目錄(存放輸入及輸出檔案的資料夾路徑)

`setwd('C://file/path')`
更改目前工作目錄

透過RStudio設定專案可指定所有檔案的工作目錄

向量

製造向量

<code>c(2, 4, 6)</code>	2 4 6	直接列出數值元素
<code>2:6</code>	2 3 4 5 6	使用數值元素製造數值
<code>seq(2, 3, by=0.5)</code>	2.0 2.5 3.0	使用函式製造連續數值
<code>rep(1:2, times=3)</code>	1 2 1 2 1 2	使用函式製造重複數值

向量運算函式

<code>sort(x)</code> 回傳排序的數值	<code>rev(x)</code> 顛倒數值順序
<code>table(x)</code> 表列數值次數	<code>unique(x)</code> 查看數值種類

索引向量內數值

位置索引

<code>x[4]</code>	第四位數值
<code>x[-4]</code>	第四位之外的數值
<code>x[2:4]</code>	第二到第四位數值
<code>x[-(2:4)]</code>	第二到第四位之外的數值
<code>x[c(1, 5)]</code>	第一位和第五位數值

數值索引

<code>x[x==10]</code>	等於10的數值
<code>x[which(x==10)]</code>	等於10的數值
<code>x[x<0]</code>	小於0的數值
<code>x[x%in%c(1, 2, 5)]</code>	符合集合{1, 2, 5}的數值

命名索引

`x['apple']` 命名是'apple'的數值

迴圈與迭代

For 迴圈

```
for (variable in sequence) {  
  Dosomething  
}
```

範例

```
for (i in 1:4) {  
  j <- i + 10  
  print(j)  
}
```

If 條件式

```
if (condition) {  
  Dosomething  
} else {  
  Dosomething  
}
```

範例

```
if (i > 3) {  
  print('Yes')  
} else {  
  print('No')  
}
```

While 迴圈

```
while (condition) {  
  Dosomething  
}
```

範例

```
while (i < 5) {  
  print(i)  
  i <- i + 1  
}
```

函式

```
func_name <- function(var) {  
  Dosomething  
  return(new_variable)  
}
```

範例

```
square <- function(x) {  
  squared <- x*x  
  return(squared)  
}
```

讀取與寫入資料

也可參考 readr 套件功能

讀取資料檔	寫入資料檔	說明
<code>df <- read.table('file.txt')</code>	<code>write.table(df, 'file.txt')</code>	讀寫格式化分欄的純文字檔
<code>df <- read.csv('file.csv')</code>	<code>write.csv(df, 'file.csv')</code>	讀寫逗號分欄的純文字檔
<code>load('file.RData')</code>	<code>save(df, file = 'file.RData')</code>	讀寫R映像檔

比對條件

<code>a == b</code>	等於	<code>a > b</code>	大於	<code>a >= b</code>	大於或等於	<code>is.na(a)</code>	等於NA
<code>a != b</code>	不等於	<code>a < b</code>	小於	<code>a <= b</code>	小於或等於	<code>is.null(a)</code>	等於null

資料型態

R的資料型態能互相轉換。表單內排序較前的數值可轉換為排序較低的數值。

as.logical	TRUE, FALSE, TRUE	邏輯值向量
as.numeric	1, 0, 1	整數或浮點數向量
as.character	'1', '0', '1'	字符及字串向量；分析資料轉成factor更好
as.factor	'1', '0', '1' levels: '1', '0'	附加level的字符向量；建立統計模型必用

數學運算函式

log(x)	自然對數	sum(x)	總和
exp(x)	自然指數	mean(x)	平均
max(x)	最大值	median(x)	中位數
min(x)	最小值	quantile(x)	百分位數
round(x, n)	顯示數值到小數點第n位	rank(x)	次序位數
signif(x, n)	小數點第n位起簡寫數值	var(x)	樣本變異數
cor(x, y)	相關。	sd(x)	樣本標準差

指派變數數值

```
> a <- 'apple'
> a
[1] 'apple'
```

R環境檢視函式

ls() 列出已指派數值的變數
rm(x) 移除變數x
rm(list = ls()) 移除全部已指派數值的變數

在RStudio · 可透過環境子視窗 (environment) 瀏覽已指派數值的變數

列表

```
l<-list(x=1:5,y=c('a','b'))
列表物件集合型態及數量不一致的資料
```

<code>l[[2]]</code>	<code>l[1]</code>	<code>l\$x</code>	<code>l['y']</code>
取得列表第二個物件數值	取得列表第一個物件名稱與數值	取得列表x物件數值	取得列表y物件名稱與數值

矩陣

<code>m[2,]</code> - 取得第2列數值	<code>t(m)</code> 轉置矩陣
<code>m[, 1]</code> - 取得第1欄數值	<code>m %*% n</code> 矩陣內積
<code>m[2, 3]</code> - 取得第2列第3欄數值	<code>solve(m,n)</code> 解聯立方程式, m必須是對角矩陣

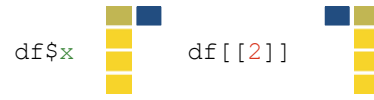
資料框架

也可參考 **dplyr** 套件功能

```
df <- data.frame(x=1:3, y=c('a','b','c'))
包含至少一筆數值數量相同的向量
```

x	y
1	a
2	b
3	c

列表式索引



預覽資料框架的函式

`view(df)` 列印全部數值
`head(df)` 列印首6列數值

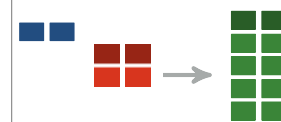
陣列式索引

<code>df[, 2]</code>	<code>nrow(df)</code> 顯示列數
<code>df[2,]</code>	<code>ncol(df)</code> 顯示欄數
<code>df[2, 2]</code>	<code>dim(df)</code> 顯示欄數與列數

cbind - 依欄合併



rbind - 依列合併



字串

也可參考 **stringr** 套件功能

<code>paste(x, y, sep = '')</code>	無縫隙拼接字串x與y
<code>paste(x, collapse = '')</code>	無縫隙拼接字串x所有數值
<code>grep(pattern, x)</code>	找出字串x符合正則表示式的數值
<code>gsub(pattern, replace, x)</code>	取代字串x內符合正則表示式的數值
<code>toupper(x)</code>	轉換成大寫字符
<code>tolower(x)</code>	轉換成小寫字符
<code>nchar(x)</code>	計算字串的字符個數

因子

factor(x) 將向量轉換成因子，並設定層次(levels)及層次順序
cut(x, breaks = 4) 指定切割區間數，將數值向量轉換成因子

統計函式

<code>lm(y ~ x, data=df)</code> 線性模型	<code>t.test(x, y)</code> 執行獨立分組平均數t檢定	<code>prop.test</code> 執行分組平均比例顯著性考驗
<code>glm(y ~ x, data=df)</code> 通用線性模型	<code>pairwise.t.test</code> 計算t檢定的相依樣本變數相關性	<code>aov</code> 變異數檢定
<code>summary()</code> 輸出線性模型分析報告		

機率分佈函式

	隨機變數	機率密度函數	累積機率函數	量數
常態分佈	rnorm	dnorm	pnorm	qnorm
卜瓦松分佈	rpois	dpois	ppois	qpois
二項分佈	rbinom	dbinom	pbinom	qbinom
均勻分佈	runif	dunif	punif	qunif

繪圖

也可參考 **ggplot2** 套件功能

<code>plot(x)</code> 依向量x數值順序繪圖	<code>plot(x, y)</code> 向量x與向量y的散佈圖	<code>hist(x)</code> 向量x的直方圖
------------------------------------	--	---------------------------------

日期

參考 **lubridate** 套件功能