

# R 軟體安裝和使用界面介紹 以及

## R 資料分析應用：描述性統計和平均數檢定

蔡靜雯 副統計分析師

### 第1部分 R軟體安裝和使用環境介紹

#### 1-1、R軟體下載和安裝

R是免費下載使用的科學與統計軟體，依以下完成步驟下載與安裝

- 下載步驟

- (1) 先到R的首頁 <https://www.r-project.org> ，
- (2) 點選左側選單 Download 抬頭下的 CRAN ，
- (3) 選取 Taiwan 下的任一個下載網址，
- (4) 然後在 Download and Install R 的格子裡選取 Download R for Windows (如果使用的系統是 Linux 或 Mac 就選取 Download R for Linux 或 Download R for (Mac) OS X ，
- (5) 接著，首次安裝點選 base ，
- (6) 再選擇 [Download R 3.5.0 for Windows](#) (目前最新版本是R 3.5.0) 即可下載R軟體。

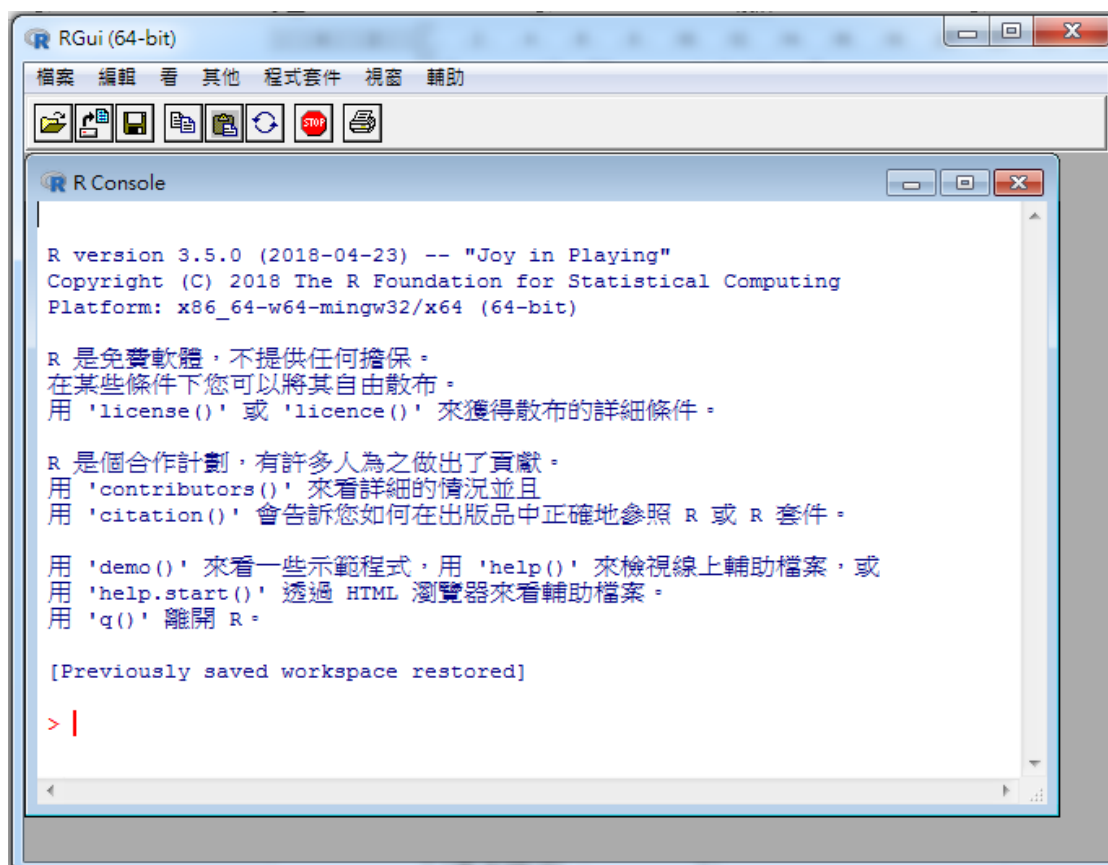
- 安裝步驟

- (1) 執行下載完成的 R-3.5.0-win.exe ，
- (2) 選擇在安裝過程中使用的語言繁體中文，按確定，

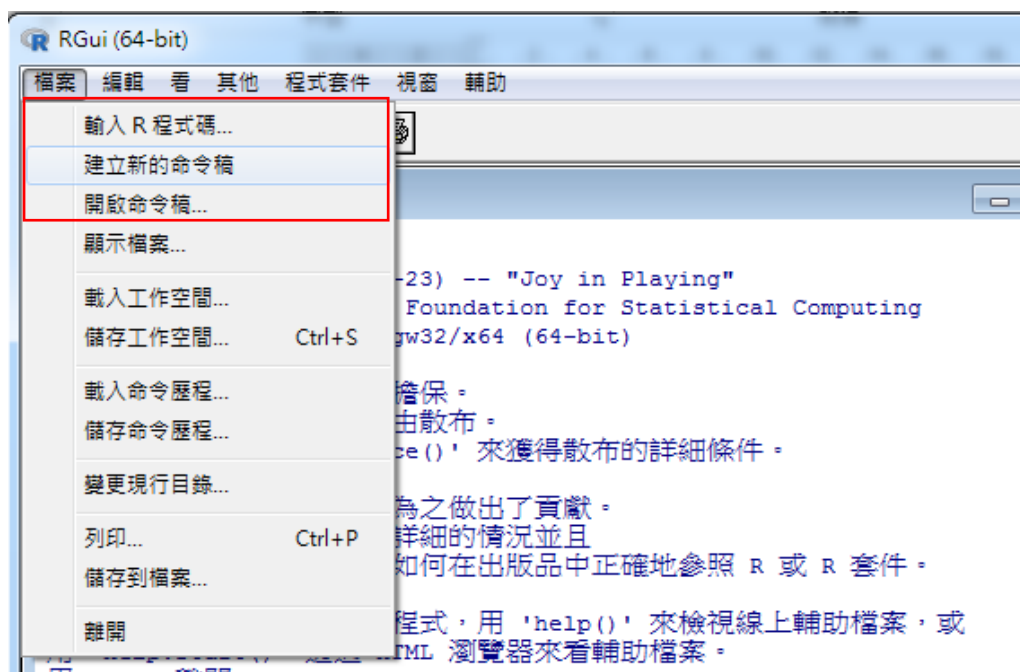
- (3) 接著按下一步，選擇安裝資料夾(預設是在C:\Program Files\R\R-3.5.0)，再按下一步，
- (4) 選擇安裝的元件(一般會自動勾選好，可依需求取消或增加選項)，再按下一步，
- (5) 點選 No (accept defaults)，再按下一步，
- (6) 選擇安裝程式建立程式的捷徑位置(預設是在C:\Program Files\R)，再按下一步，
- (7) 選擇要執行附加工作(一般會自動勾選好，可依需求取消或增加選項)，再按下一步，就會執行安裝，
- (8) 安裝完成後，點選完成，即可在桌面或開始目錄點開 R 軟體使用，

## 1-2、使用界面介紹

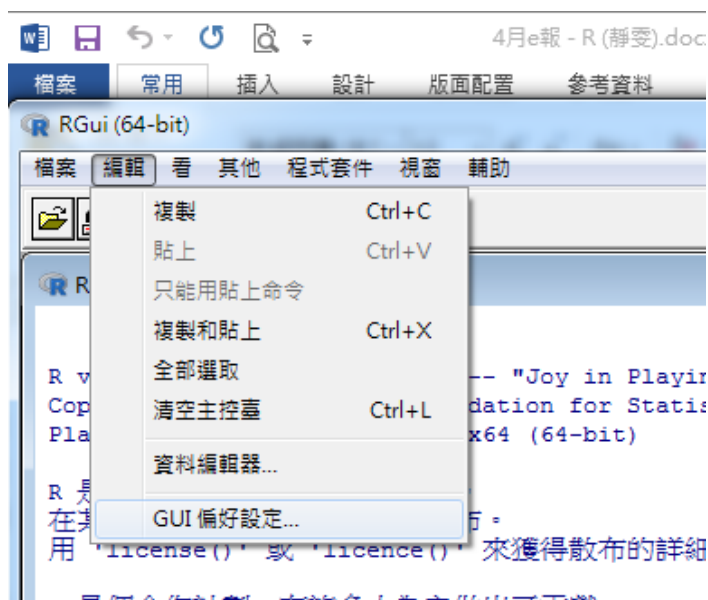
R的Windows圖形界面模式(GUI)：



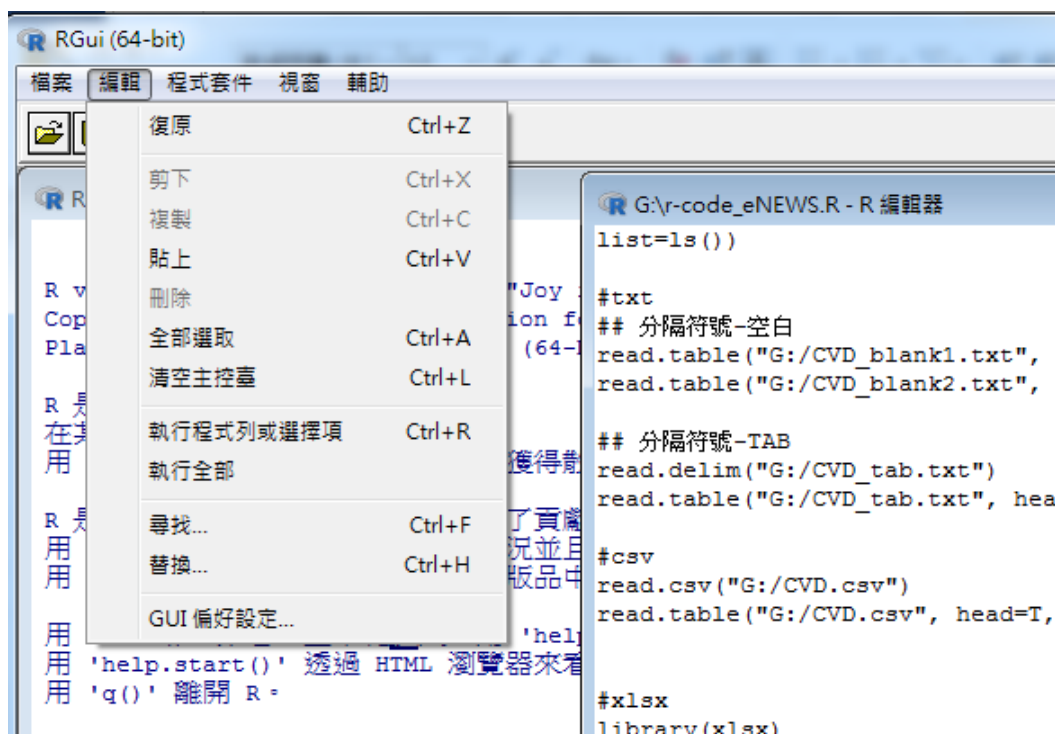
**檔案選單：**可以點選「輸入R程式碼」執行事先寫好的R程式檔，或點選「建立新的命令稿」撰寫新程式，或是「開啟命令稿」開啟外部已寫好的R程式檔。



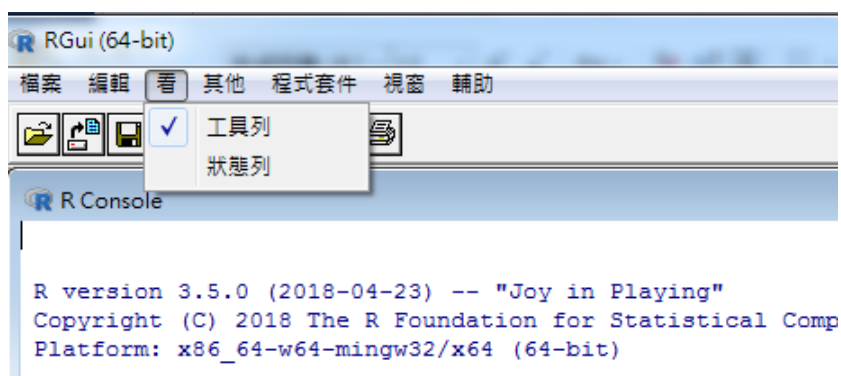
**編輯選單：**滑鼠點選圖形界面模式(GUI)，再點選編輯選單，選單裡的選項是可在GUI執行的動作或設定GUI的使用偏好



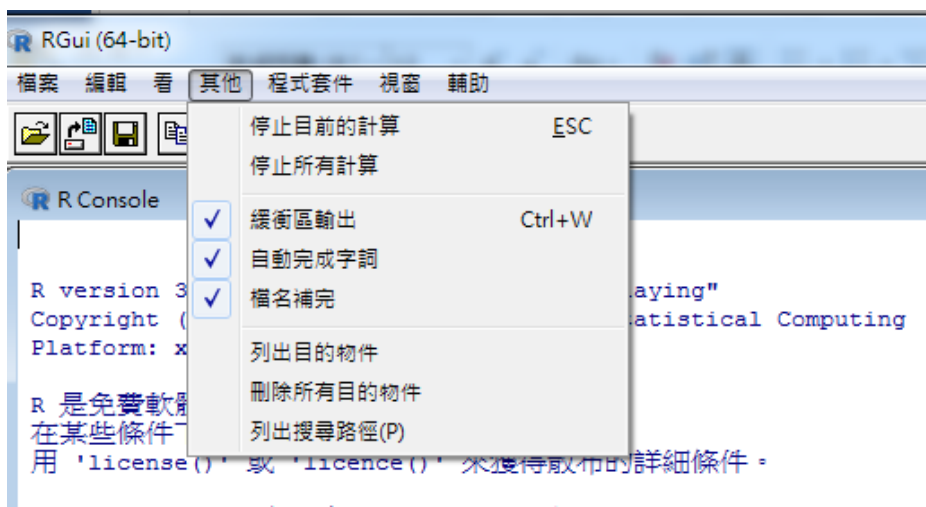
**編輯選單：**滑鼠點選在R編輯器，再點選編輯選單，選單裡的選項會增加4個可在編輯器裡執行的動作：「執行程式列或選項列Ctrl+R」、「執行全部」、「尋找」和「替換」的選項。



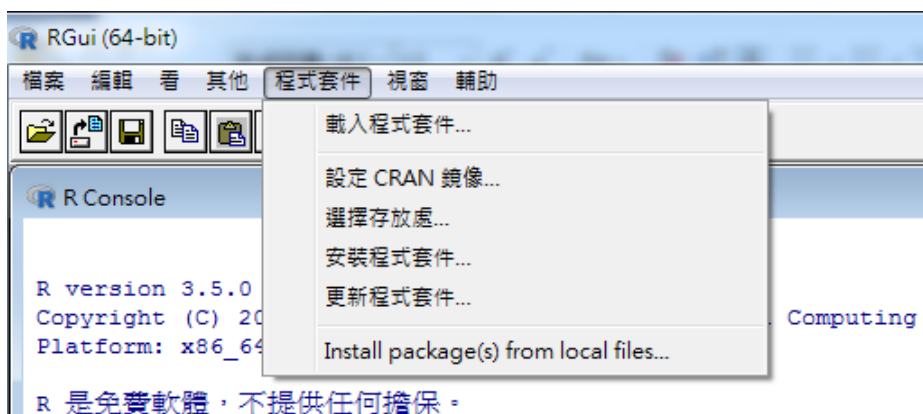
**看(檢視)選單：**選擇是否顯示工具列、狀態列(預設是只勾選顯示工具列)。



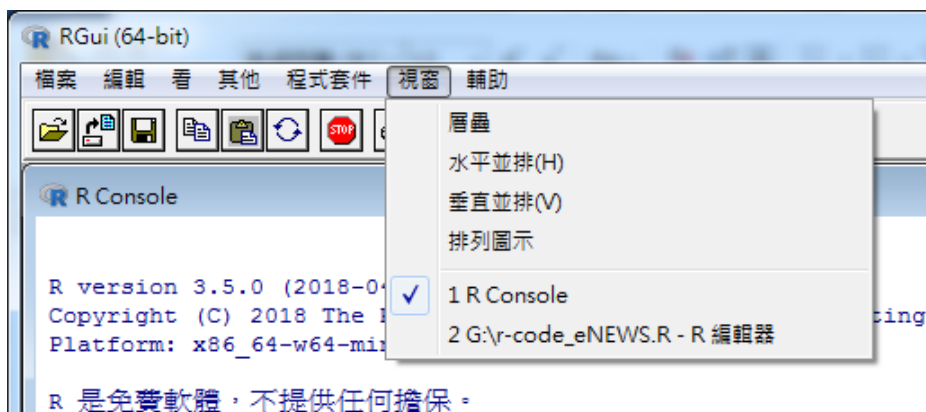
**其他選單：**如果要中止執行程式，可在這個選單裡點選「停止目前的計算 ESC」或「停止所有計算」。



**程式套件選單：**「安裝程式套件」或「更新程式套件」可在這裡點選。



**視窗選單：**GUI界面中所有視窗的顯示和排列方式。



**輔助選單：**可在這裡查詢 R 軟體指令的相關資訊。



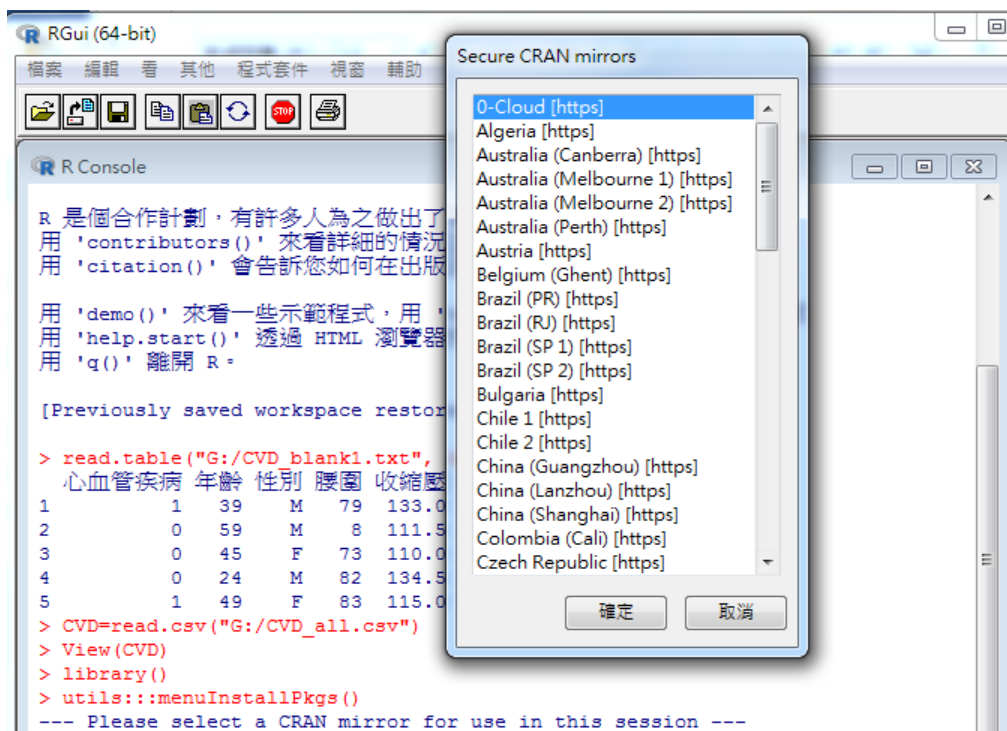
### 3、安裝程式套件

R軟體安裝完成後，安裝目錄中會有基本套件函數，不需再用 **library(套件名稱)** 函數指令載入套件。若要查詢已經安裝的套件，可直接輸入 **library()** 查詢。

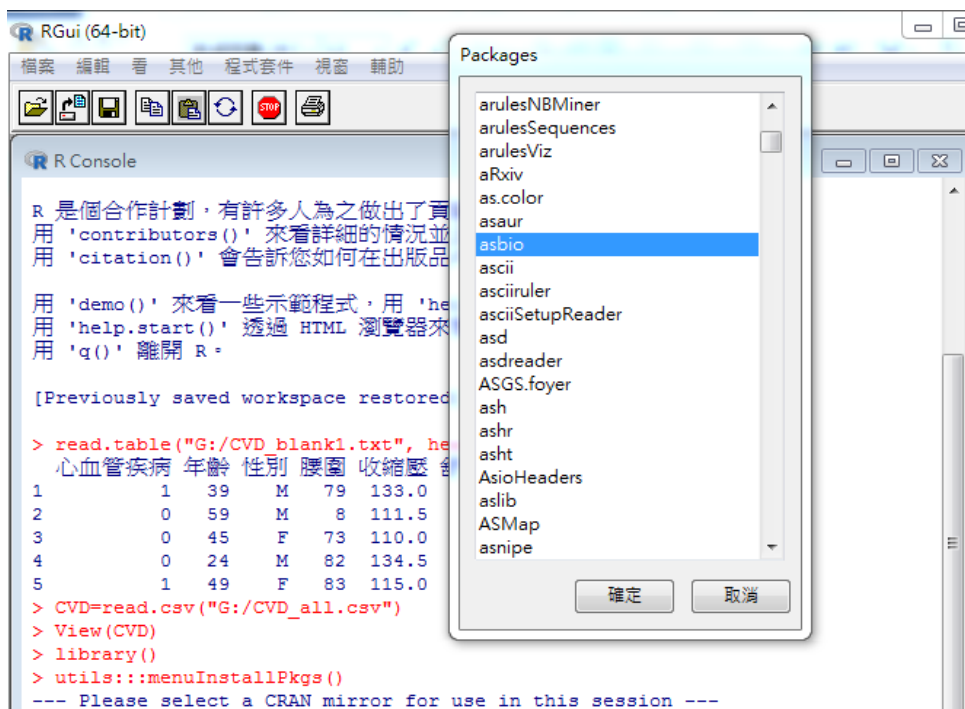
若要使用安裝目錄中沒有的套件，可使用兩種方式安裝

✓ 圖形界面(GUI)點選安裝：

- (1) 點選程式套件選單中的「安裝程式套件」，會出現Secure CRAN mirrors的選單，任意選擇一個安裝套件的來源鏡射網站，按確定。



- (2) 會再出現 Packages 選單，選取欲安裝的套件名稱，按確定，即可完成套件安裝。要使用套件裡的函數前，記得先用 **library(套件名稱)** 函數指令載入套件。



- ✓ 執行程式函數指令安裝：**install.packages("套件名稱")**

## 第2部分 R軟體的資料分析應用

### 1、讀入資料

R軟體讀取外部資料有多種函數的語法可使用，以下將介紹三種外部資料格式的匯入方式：

#### 1-1 讀取txt格式的檔案

##### **read.table** 函數

###### 【基本語法】

```
read.table(file, header=F, sep="", na.strings = "NA", blank.lines.skip=T)
```

參數說明：

file 讀入檔案的路徑和檔案名稱、格式

header 檔案第一列是否為變數名稱，預設為FALSE(否)

sep 指定檔案的分隔符號，預設為空格""

na.strings 指定檔案中的遺失值以什麼字串來標示，預設為"NA"

blank.lines.skip 檔案中若有空白資料欄位是否跳過，預設為TRUE(是)

#### ➤ 範例1-1.1：

若 G:\CVD\_blank1.txt 檔案中無空白欄位，檔案路徑為G:\且以空格分隔的文字檔

```
心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1 39 M 79 133 72 1
0 59 M 8 111.5 74.5 0
0 45 F 73 110 73.5 0
0 24 M 82 134.5 80.5 1
1 49 F 83 115 84.5 0
```



## 程式碼1-1.1

```
read.table("G:/CVD_blank1.txt", head=T)
```

[說明] G:\CVD\_blank1.txt 檔案第一行為變數名稱，因此要指定 head=TRUE，TRUE要大寫，也可寫成head=T

output

```
> read.table("G:/CVD_blank1.txt", head=T)
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79   133.0   72.0    1
2           0   59    M    8   111.5   74.5    0
3           0   45    F   73   110.0   73.5    0
4           0   24    M   82   134.5   80.5    1
5           1   49    F   83   115.0   84.5    0
```

## ➤ 範例1-1.2：

若 G:\CVD\_blank2.txt 檔案中有空白欄位，檔案路徑為G:\且以空格分隔的文字檔

```
心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1 39 M 79 133 72
0 59 M 8 111.5 74.5 0
0 45 F 73 110 73.5 0
0 24 M 82 134.5 80.5 1
1 49 F 83 115 84.5 0
```

## 程式碼1-1.2

```
read.table("G:/CVD_blank2.txt", head=T, blank.lines.skip=F)
```

[說明] 檔案中有空白欄位時，需指定 blank.lines.skip=F，將空白欄位視為遺失值讀入，否則會有Error出現無法讀入資料

output

```
> read.table("G:/CVD_blank2.txt", head=T, blank.lines.skip=F)
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79   133.0   72.0   NA
2           0   59    M    8   111.5   74.5    0
3           0   45    F   73   110.0   73.5    0
4           0   24    M   82   134.5   80.5    1
5           1   49    F   83   115.0   84.5    0
```

## ➤ 範例1-1.3：

若 G:\CVD\_tab.txt 的檔案路徑為G:\且以 TAB 分隔的文字檔

心血管疾病	年齡	性別	腰圍	收縮壓	舒張壓	飲酒
1	39	M	79	133	72	1
0	59	M	78	111.5	74.5	
0	45	F	73	110	73.5	0
0	24	M	82	134.5	80.5	1
1	49	F	83	115	84.5	0

## 程式碼1-1.3

```
read.table("CVD_tab.txt", head=T, sep="\t")
```

[說明] 分隔符號為TAB時，需指定sep="\t"

[註] 當分隔符號為TAB，不管檔案中有無空白欄位，都無需指定blank.lines.skip為TRUE或FALSE，自動會將空白欄位視為遺失值讀入

## output

```
> read.table("G:/CVD_tab.txt", head=T, sep="\t")
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79   133.0   72.0    1
2           0   59    M   78   111.5   74.5   NA
3           0   45    F   73   110.0   73.5    0
4           0   24    M   82   134.5   80.5    1
5           1   49    F   83   115.0   84.5    0
```

**read.delim** 函數，範例1-1.3 可使用另一種語法將檔案讀入函數

## 【基本語法】

```
read.delim(file, head=T, sep="\t", ...)
```

[說明] read.delim 函數的預設分隔符號為"TAB"、讀入檔案第一列為變數名稱，當檔案是以TAB分隔且首列為變數名稱時，無需再指定參數，使用預設值即可，檔案中若有空白欄位會視為遺失值讀入

## 程式碼1-1.4

```
read.delim("G:/CVD_tab.txt")
```

output

```
> read.delim("G:/CVD_tab.txt")
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79  133.0   72.0    1
2           0   59    M   78  111.5   74.5   NA
3           0   45    F   73  110.0   73.5    0
4           0   24    M   82  134.5   80.5    1
5           1   49    F   83  115.0   84.5    0
```

## 1-2 讀取 CSV 格式的檔案

### 【基本語法】

```
read.csv(file, head=T, sep="," , ...)
```

[說明] read.csv 函數的預設分隔符號為逗點","、讀入檔案第一列為變數名稱，當檔案是以逗點分隔且首列為變數名稱時，無需再指定參數，使用預設值即可，檔案中若有空白欄位會視為遺失值讀入

### ➤ 範例1-2

若 G:\CVD.csv 為 CSV 格式的檔案

	A	B	C	D	E	F	G
1	心血管疾病	年齡	性別	腰圍	收縮壓	舒張壓	飲酒
2	1	39	M	79	133	72	1
3	0	59	M	78	111.5	74.5	
4	0	45	F	73	110	73.5	0
5	0	24	M	82	134.5	80.5	1
6	1	49	F	83	115	84.5	0
7							

### 程式碼1-2.1

```
read.csv("G:/CVD.csv")
```

output

```
> read.csv("G:/CVD.csv")
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79  133.0   72.0    1
2           0   59    M   78  111.5   74.5   NA
3           0   45    F   73  110.0   73.5    0
4           0   24    M   82  134.5   80.5    1
5           1   49    F   83  115.0   84.5    0
```

CSV檔案也可使用 **read.table** 函數將檔案讀入

程式碼1-2.2

```
read.table("G:/CVD.csv", head=T, sep=",")
```

output

```
> read.table("G:/CVD.csv", head=T, sep=",")
  心血管疾病 年齡 性別 腰圍 收縮壓 舒張壓 飲酒
1           1   39    M   79   133.0   72.0    1
2           0   59    M   78   111.5   74.5   NA
3           0   45    F   73   110.0   73.5    0
4           0   24    M   82   134.5   80.5    1
5           1   49    F   83   115.0   84.5    0
```

1-3 讀取 xlsx 格式的檔案

**read.xlsx** 函數

【基本語法】

```
read.xlsx(file, sheetIndex, header=TRUE, encoding="unknown", ...)
```

參數說明：

file 讀入檔案的路徑和檔案名稱、格式

sheetIndex 指定要讀入檔案中的第幾張分頁

header 檔案第一列是否為變數名稱，預設為是(TRUE)

encoding 檔案中文字字串的編碼

➤ 範例1-3

若 G:\CVD.xlsx 為 xlsx 格式的檔案

	A	B	C	D	E	F	G
1	心血管疾病	年齡	性別	腰圍	收縮壓	舒張壓	飲酒
2	1	39	M	79	133	72	1
3	0	59	M	78	111.5	74.5	
4	0	45	F	73	110	73.5	0
5	0	24	M	82	134.5	80.5	1
6	1	49	F	83	115	84.5	0
7							
8							

## 程式碼1-2.1

```
library(xlsx)
```

```
read.xlsx("G:/CVD.xlsx", sheetIndex=1)
```

若出現Error無法順利讀入資料時，可試著指定參數encoding="UTF-8"

```
read.xlsx("G:/CVD.xlsx", sheetIndex=1, encoding="UTF-8")
```

[說明] read.xlsx 函數需先安裝程式套件：xlsx(安裝方式請參考第一部分3、安裝程式套件)，安裝完成後，用library函數匯入 xlsx 套件即可。

[註] 讀入 xlsx 格式的檔案，容易出現編碼問題而無法讀入資料，建議將檔案另存為 csv 檔，再讀入檔案

## output

```
> read.xlsx("G:/CVD.xlsx", sheetIndex=1, encoding="UTF-8")
```

	心血管疾病	年齡	性別	腰圍	收縮壓	舒張壓	飲酒
1	1	39	M	79	133.0	72.0	1
2	0	59	M	78	111.5	74.5	NA
3	0	45	F	73	110.0	73.5	0
4	0	24	M	82	134.5	80.5	1
5	1	49	F	83	115.0	84.5	0

## 2、描述性統計

## 描述性統計的相關函數1

**mean(x, ...)** 平均數，其中x為放入的資料

**median(x, ...)** 中位數

**min(x, ...)** 最小值

**max(x, ...)** 最大值

**quantile(x, ...)** 百分位數

**summary(object, ...)** 敘述統計量，包含最小值、第25百分位數、中位數、平均數、第75百分位數和最大值

## 描述性統計的相關函數2

**var(x, ...)** 變異數

**sd(x, ...)** 標準差

**sum(x, ...)** 總和

**range(x, ...)** 全距

**cv(x, ...)** 變異系數

**skewness(x, ...)** 偏態系數，需下載套件e1071

**kurtosis(x, ...)** 峰度系數，需下載套件e1071

### ➤ 範例資料檔案

範例資料檔案可從臺北醫學大學管理學院的生物統計研究中心網頁中的生統課程資料檔位置下載 <http://biostat.tmu.edu.tw/index.php/download>，下載檔案名稱 "CVD\_All" 為基隆地區整合篩檢計畫(KCIS)這個樣本收集資料，其檔案格式為CSV檔，檔案詳細內容請參考生統課程資料檔中的資料說明檔。

### ➤ 範例2

欲了解基隆地區整合篩檢計畫這個樣本收集資料中的平均年齡

### 程式碼2

```
#讀入資料，命名為CVD
CVD=read.csv("G:/CVD_all.csv")

#步驟2 偵測資料中是有遺失值
na.fail(CVD)

#步驟3 若資料中有遺失值，要先去除遺失值
CVD_ok=na.omit(CVD)

#步驟4
mean(CVD_ok$年齡)
```

output

```
> na.fail(CVD)
Error in na.fail.default(CVD) : missing values in object
> CVD_ok=na.omit(CVD)
> mean(CVD_ok$年齡)
[1] 46.42333
```

當資料中有遺失值時，有些函數指令無法計算，在這種情況下，就要先用 **na.omit(資料檔案)** 函數把資料中的遺失值去除，再執行計算。計算結果，基隆地區整合篩檢計畫這個資料中的平均年齡為46.42333歲。其他的描述性統計函數的使用方式，以此類推。

### 3、平均數檢定

**t.test** 函數

#### 【基本語法】

```
t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0,
       paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)
```

參數說明：

x 數值資料，不可為空

y 數值資料，可放可不放，預設為 NULL (不放)

alternative 設定雙尾、左尾或右尾檢定，預設為雙尾檢定

mu 若為單一樣本平均數檢定，mu為設定檢定平均數，預設為0

若為雙樣本平均數檢定，mu為設定檢定平均數值差異，預設為0

paired 設定是否為成對樣本，預設為 FALSE (否)

var.equal 設定兩樣本的變異數是否相等，預設為 FALSE (否)

conf.level 設定信賴區間，預設為 0.95 (95%信賴區間)

(1) 單一樣本平均數 t 檢定

## ➤ 範例3-1

腰圍能反映腹部肥胖的多寡，藉以反映內臟脂肪的堆積，根據衛生福利部國民健康署的測量標準，若男性腰圍 $\geq 90$ 公分，女性腰圍 $\geq 80$ 公分，表示腹部肥胖，想了解基隆地區整合篩檢計畫(KCIS)這個樣本收集資料中，男性平均腰圍是否 $< 90$ 公分？

## 程式碼3-1

```
#讀入資料，命名為CVD
CVD=read.csv("G:/CVD_all.csv")

#步驟1. 切割資料，以找出男性的資料並命名為CVD_m
CVD_m=subset(CVD, 性別==1)    #[補充1]

#步驟2. 單一樣本 t 檢定分析：男性平均腰圍是否<90公分為左尾檢定
t.test(CVD_m$腰圍, alternative="less", mu=90)    #[補充2]
```

output - 單一樣本 t 檢定分析

```
> t.test(CVD_m$腰圍, alternative="less", mu=90)

One Sample t-test

data:  CVD_m$腰圍
t = -91.209, df = 23416, p-value < 2.2e-16
alternative hypothesis: true mean is less than 90
95 percent confidence interval:
 -Inf 84.47406
sample estimates:
mean of x
 84.37257
```

## [分析結果]

此範例是看男性平均腰圍是否 $< 90$ 公分，為左尾檢定，虛無假設為男性平均腰圍 $\geq 90$ 公分，對立假設為男性平均腰圍 $< 90$ 公分，根據output的分析結果，P值 $< 2.2e-16$ 拒絕虛無假設，代表基隆地區整合篩檢計畫樣本收集資料中的男性平均腰圍，在統計上沒有顯著超過90公分的測量標準。



(2) 獨立雙樣本平均數差異 t 檢定

## ➤ 範例3-2

美國疾病管制局(CDC)出版的菸害報告亦指出，男女性吸菸者比非吸菸者有高達2~7倍的心血管疾病死亡風險，而肥胖，尤其是腹部肥胖則會增加心血管疾病發生的機會，進而懷疑，抽菸者和沒有抽菸者的腰圍是否會有差異？

## 程式碼3-2

```
#步驟1. 將有、無抽菸者的資料分開存檔，命名為 smoke_yes 和 smoke_no
smoke_yes=subset(CVD, 抽菸==1)
smoke_no=subset(CVD, 抽菸==0)

#步驟2. 雙樣本變異數檢定：有無抽菸者的腰圍變異數是否相等
var.test(smoke_yes$腰圍,smoke_no$腰圍, alternative="two.sided")

#步驟3. 獨立雙樣本平均數差異 t 檢定：有無抽菸者的平均腰圍差是否等於0
t.test(smoke_yes$腰圍, smoke_no$腰圍, mu=0, var.equal=F)

#[註] t.test 函數預設alternative="two.sided"，若不指定就會使用預值
```

## output - 雙樣本變異數檢定

```
> var.test(smoke_yes$腰圍,smoke_no$腰圍, alternative="two.sided" )

      F test to compare two variances

data:  smoke_yes$腰圍 and smoke_no$腰圍
F = 1.0282, num df = 17279, denom df = 45226, p-value = 0.02736
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.003108 1.054096
sample estimates:
ratio of variances
 1.028215
```

在執行獨立雙樣本平均數差異 t 檢定前，需先檢定有、無抽菸者腰圍的變異數是否相等，根據上表的output結果，P值為 $0.02736 < \alpha = 0.05$ ，拒絕虛無假設：有、無抽菸者腰圍的變異數比例等於1，即為兩變異數不相等。因此，下一步驟的獨立雙樣本平均數差異 t 檢定中，其變異數是否相等的參數設定為 `var.equal= FALSE`。

output - 獨立雙樣本平均數差異 t 檢定

```
> t.test(smoke_yes$腰圍, smoke_no$腰圍, mu=0, var.equal=FALSE)

Welch Two Sample t-test

data:  smoke_yes$腰圍 and smoke_no$腰圍
t = 60.139, df = 30877, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5.429297 5.795124
sample estimates:
mean of x mean of y
 82.40398  76.79177
```

[分析結果]

比較抽菸者的腰圍是否有差異為雙尾檢定，虛無假設是抽菸者和沒有抽菸者的平均腰圍差等於0、對立假設則是抽菸者和沒有抽菸者的平均腰圍差不等於0，從output的結果顯示， $P$ 值 $<2.2e-16$  拒絕虛無假設，即表示基隆地區整合篩檢計畫這筆資料中，有、無抽菸者的平均腰圍在統計上有顯著差異。

(3)成對雙樣本平均數差異 t 檢定

#### ➤ 範例3-3

一般人常誤以為，血壓只要在正常範圍內，就不會罹患心血管疾病。有研究指出，如果收縮壓和舒張壓相差超過60，表示血管彈性變差，即使血壓正常，仍存在血管硬化，因而增加罹病的風險。藉此，同樣使用基隆地區整合篩檢計畫的資料，來檢定研究對象的平均脈壓差是否超過標準值( $>60$  mmHg)？

程式碼3-3

```
#步驟1 雙樣本變異數檢定：收縮壓和舒張壓的變異數是否相等
var.test(CVD$收縮壓, CVD$舒張壓, alternative="two.sided")

#步驟2. 成對雙樣本平均數差異 t 檢定：脈壓差是否>60為右尾檢定
t.test(CVD$收縮壓, CVD$舒張壓, alternative="greater", mu=60, var.equal=F)
```

output - 雙樣本變異數檢定

```
> var.test(CVD$收縮壓, CVD$舒張壓, alternative="two.sided")

F test to compare two variances

data: CVD$收縮壓 and CVD$舒張壓
F = 3.0233, num df = 63255, denom df = 63244, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.976526 3.070776
sample estimates:
ratio of variances
      3.023284
```

同樣的在執行成對雙樣本平均數差異 t 檢定前，需先檢定兩樣本的變異數是否相等，根據output結果，P值<2.2e-16，拒絕虛無假設：收縮壓和舒張壓的變異數比例等於1，即為收縮壓和舒張壓的變異數不相等。因此，下一步驟的成對雙樣本平均數差異 t 檢定中，其變異數是否相等的參數設定為 var.equal= FALSE。

output - 成對雙樣本平均數差異 t 檢定

```
> t.test(CVD$收縮壓, CVD$舒張壓, alternative="greater", mu=60,
+        var.equal=F)

Welch Two Sample t-test

data: CVD$收縮壓 and CVD$舒張壓
t = -154.99, df = 100980, p-value = 1
alternative hypothesis: true difference in means is greater than 60
95 percent confidence interval:
 45.03943      Inf
sample estimates:
mean of x mean of y
123.26660  78.07007
```

[分析結果]

平均脈壓差是否超過標準值(>60 mmHg)為右尾檢定，虛無假設是平均脈壓差 $\leq 60$ 、對立假設則是平均脈壓差>60，從output的結果顯示，P值為 $1 > \alpha = 0.05$  不拒絕虛無假設，即表示基隆地區整合篩檢計畫參與者的脈壓差，在統計上沒有顯著超過標準值(>60 mmHg)。

(4) 獨立多樣本平均數差異檢定

## ➤ 範例3-4.1

從範例3-2已知基隆地區整合篩檢計畫這筆資料中，有、無抽菸者的平均腰圍在統計上有顯著差異，進一步想了解抽菸量的多寡其平均腰圍是否有差異？這筆資料將有抽菸者的抽菸量分為三個等級(1：每日菸量1包；2：每日菸量2包；3：每日菸量起過2包)

## 程式碼3-4.1

```
#步驟1 切割資料，以找出有抽菸者的資料並命名為CVD_somke
CVD_smoke=subset(CVD, 抽菸==1)

#步驟2 獨立多樣本平均數差異檢定
anova(lm(腰圍~as.factor(抽菸量), data=CVD_smoke))
```

## output

```
> anova(lm(腰圍~as.factor(抽菸量), data=CVD_smoke))
Analysis of Variance Table

Response: 腰圍
          Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(抽菸量)      2   21332  10666.1    99.011 < 2.2e-16 ***
Residuals            15940  1717163    107.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## [分析結果]

此範例的虛無假設：三種等級抽菸量的平均腰圍都相同，對立假設：三種等級抽菸量的抽菸者其腰圍不全相同，從output的結果顯示，P值<2.2e-16 拒絕虛無假設，即表示基隆地區整合篩檢計畫參與者中，不同等級抽菸量的抽菸者其腰圍不全相同。

## ➤ 範例3-4.2

沿續上一個範例，已知三種等級抽菸量的抽菸者其腰圍不全相同，若進一步想知道哪些抽菸量等級之間的腰圍是有差異的，就可以使用多重比較來做事後檢定，事後檢定有多種多方法，**pairw.anova** 函數中提供了五種檢定方法，分別為Fisher's Least Significant Difference(LSD)、Bonferroni test、Tukey's test、Scheffe test(雪費檢定)、Dunnett's test，各有優缺點，這邊就不多介紹。

**pairw.anova** {asbio} 函數

## 【基本語法】

```
pairw.anova(y, x, conf.level = 0.95, method = "tukey", MSE = NULL,  
df.err = NULL, control = NULL)
```

參數說明：

y 數值型資料

x 類別型的資料

conf.level 設定信賴區間，預設為 0.95 (95%信賴區間)

method 指定檢定方法，有5種檢定方法可選擇"lsd"、"bonf"、 "tukey"、  
"scheffe"、"dunnett"，預設為"tukey"

MSE ANOVA模型中的MSE值，預設為NULL

df.err ANOVA模型中的自由度差，預設為NULL

control 當使用Dunnett's test 時，需指定對照組

## 程式碼3-4.2

```

#步驟1 偵測資料中是有遺失值
na.fail(CVD_smoke)

#步驟2 若資料中有遺失值，要先去除遺失值
CVD_smoke.ok=na.omit(CVD_smoke)

#步驟3 多重比較
library(asbio)
# LSD test
pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量), method="lsd")

# Bonferroni test
pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量), method="bonf")

# Tukey's test
pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量), method="tukey")

# Scheffe test(雪費檢定)
pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
            method="scheffe")

# Dunnett's test 需指定對照組，這邊指定對照組為1(每日菸量1包)
pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
            method="dunnett", control=1)

```

output - 偵測資料中是有遺失值

```

> na.fail(CVD_smoke)
Error in na.fail.default(CVD_smoke) : missing values in object

```

根據output顯示，CVD\_smoke資料中有遺失值存在。**pairw.anova** 函數中，若有遺失值無法顯示完整的檢定數值，因此，要多加一個步驟去除遺失值，再執行多重比較檢定。

output

### Fisher's Least Significant Difference(LSD)

```
> # LSD test
> pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
+             method="lsd")

95% LSD confidence intervals
```

	LSD	Diff	Lower	Upper	Decision	Adj. p-value
mu1-mu2	0.56175	-3.31908	-3.88083	-2.75732	Reject H0	0
mu1-mu3	1.74878	-4.90434	-6.65312	-3.15555	Reject H0	0
mu2-mu3	1.81957	-1.58526	-3.40483	0.23432	FTR H0	0.08771

### Bonferroni test

```
> # Bonferroni test
> pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
+             method="bonf")

95% Bonferroni confidence intervals
```

	Diff	Lower	Upper	Decision	Adj. p-value
mu1-mu2	-3.31908	-4.00525	-2.63291	Reject H0	0
mu1-mu3	-4.90434	-7.04044	-2.76823	Reject H0	0
mu2-mu3	-1.58526	-3.80783	0.63732	FTR H0	0.263135

### Tukey's test

```
> # Tukey's test
> pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
+             method="tukey")

95% Tukey-Kramer confidence intervals
```

	Diff	Lower	Upper	Decision	Adj. p-value
mu1-mu2	-3.31908	-3.99083	-2.64733	Reject H0	0
mu1-mu3	-4.90434	-6.99555	-2.81312	Reject H0	0
mu2-mu3	-1.58526	-3.76112	0.59061	FTR H0	0.202279

## Scheffe test(雪費檢定)

```
> # Scheffe test(雪費檢定)
> pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
+             method="scheffe")
```

95% Scheffe confidence intervals

	Diff	Lower	Upper	Decision	Adj. P-value
mu1-mu2	-3.31908	-4.02065	-2.6175	Reject H0	0
mu1-mu3	-4.90434	-7.08839	-2.72028	Reject H0	0
mu2-mu3	-1.58526	-3.85772	0.68721	FTR H0	0.232704

## Dunnett's test

```
> # Dunnett's test 需指定對照組，這邊指定對照組為1(每日菸量1包)
> pairw.anova(CVD_smoke.ok$腰圍, as.factor(CVD_smoke.ok$抽菸量),
+             method="dunnett", control=1)
```

95% Dunnett confidence intervals

	Diff	Lower	Upper	Decision
mu2-mu1	3.319078	2.677529	3.960628	Reject H0
mu3-mu1	4.904335	2.907145	6.901525	Reject H0

## [分析結果]

不管是用哪種事後檢定法，都可以從output中的Decision來判斷兩個類別間是否有差異，這邊的虛無假設是兩種不同級別抽菸量的抽菸者腰圍相同，對立假設則為兩種不同級別抽菸量的抽菸者腰圍不同。以LSD的檢定結果來看，mu1(每日菸量1包)和mu2(每日菸量2包)的Decision為Reject H0，即表示每日菸量1包和每日菸量2包的抽菸者腰圍在統計上有顯著差異。



## 補充說明

### [補充1] **subset** 函數

#### 【基本語法】

**subset(x, subset, select, ...)**

參數說明：

x 將被切割的資料

subset 條件判斷式，要保留下來的類別或範圍

select 挑選要留下的變數，若未指定，預設為檔案中的變數都保留

### [補充2] 檔案名稱\$變數名稱

是表示抓出檔案中的某個變數資料，和中文中“的”的意思相同，CVD\_m\$腰圍即CVD\_m檔案的腰圍變數資料

## 參考資料

1. R 軟體 應用統計方法 陳景祥編著 東華書局
2. 現代統計學 二版 吳忠武等著 華泰文化書局