

初學 語言的 60 分鐘

廖鎮磐 <andrew.43@gmail.com>

東海大學生命科學系

2015 年台灣生態研究網年會

2015 年 3 月 14 日於蓮華池研究中心



© 2015 廖鎮磐 (Chen-Pan Liao)

本文件採用姓名標示-相同方式分享 4.0 國際授權 (CC BY-SA 4.0)。¹

歡迎下載本投影片及資料檔案。²

¹ 詳情請見 http://creativecommons.org/licenses/by-sa/4.0/deed.zh_TW。

² https://www.dropbox.com/sh/7p9cluglptyk25a/AAAmPWrSMVf_tC9-cL4JSscCa?dl=0

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

今天主題

目標

- 不怕使用 R 這類以文字指令進行的工作方式。
- 如何自己救自己。
- 如何請別人救自己。
- 實作一些常見的統計分析與繪圖。

預設聽眾

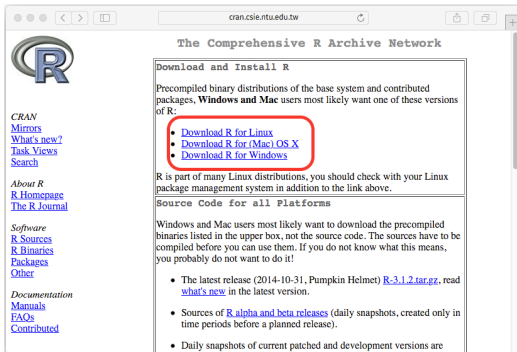
- 修過至少 3 學分的統計學。
- 從沒使用過 R 或其它統計軟體。
- 從沒學過任何程式語言。

R 的特色

- 自由、免費、跨平台。
- 是一種「程式語言」，像 Python、Perl、JAVA 等。
- 是一種「統計工具」，像 SAS、SPSS 等。
- 強大的視覺化工具，畫專業的圖，但需要經驗。
- 套件豐富，不同自己重新寫程式。

安裝 R 語言

- 1 到達 <http://www.r-project.org/>
- 2 點選 Download, Packages (CRAN)
- 3 選擇作業平台



選用適當的 R 程式編輯器

- 建議以純文字編輯器撰寫 R 程式碼，並儲存成「.R」檔。
- 「語法多色支援」、「語法提示」、「即時執行」等功能，增加撰寫效率。
- 支援 R 語言的編輯器很多，有興趣請自行試用。

RStudio 強大、整合性高、專為 R 程式員設計。³

Notepad++ 老字號的純文字編輯器，有和 R 相配合的外掛。⁴

Atom 與 GitHub 配合度高。⁵

³ <http://www.rstudio.com/>

⁴ <http://notepad-plus-plus.org/>

⁵ <https://atom.io/>

初次見面：R 是計算機

```
> 2.4 + 42
```

```
[1] 44.4
```

```
> 4 ^ 2
```

```
[1] 16
```

```
> sqrt(100)
```

```
[1] 10
```

```
> 100 ^ 0.5
```

```
[1] 10
```

```
> a <- 1
```

```
> a
```

```
[1] 1
```

```
> 1 -> b
```

```
> b
```

```
[1] 1
```

```
> a + b
```

```
[1] 2
```


大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

什麼是程式語言的函數

使用某函數的語法通則

函數名(第一引數名 = 某值, 第二引數名 = 某值, ...)

- 程式語言的函數 (function) 提供一個特定的功能，可以有引數（輸入值）和回傳值（輸出值）。
- 操作 R 的過程，幾乎就是使用各種 function 的過程。
- 若需要進行某一套程式多次，可以將之改寫成函數以利重覆使用。
- 試試看 `seq(from = 0, to = 9)` 的回傳值是什麼？
- 用中文說明上面的程式：「在 `seq()` 這個 function 中，第一個引數名為 `from`，表示起始值，其值為 0；第二個引數名為 `to`，表示終點值，其值是 9。」

函數的使用手冊

- 觀看某個函數的使用手冊：`?函數名` 或 `help(函數名)`。
- 請看看 `?seq`。
- 使用手冊中都有以下資訊：

Description 函數的功能。

Usage 基本語法，包括了引數的順序和預設值。

Arguments 引數的細節。

Details 函數的詳細內容。

Value 回傳值的內容。

See Also 其它相關的函數。

Examples 使用範例。

引數的預設值

seq() 的基本語法

```
seq(from = 1, to = 1, ...)
```

- 在使用手冊中可以看出：
第一個引數 `from` 的預設值是 1。
第一個引數 `to` 的預設值是 1。
- 使用者未定義時採用的值，就是預設值。
- 方便快捷使用。
- 例如：
`seq(from = 10)` 和
`seq(from = 10, to = 1)` 是相等的。

引數的順序

seq() 的基本語法

```
seq(from = 1, to = 1, ...)
```

- 當明確指定引數名時，引數的順序無所謂。例如：
`seq(from = 0, to = 9)` 和
`seq(to = 9, from = 0)` 同義。
- 當引數的順序與該函數要求的順序相同時，可以省略引數名。例如：
`seq(from = 0, to = 9)` 可以省略為
`seq(0, 9)` 的形式。

引數的綜合練習

seq() 的基本語法

```
seq(from = 1, to = 1, ...)
```

試回答下列程式的回傳值為何？

- `seq(from = 3, to = 1)`
- `seq(3, to = 1)`
- `seq(from = 3, 1)`
- `seq(3, 1)`
- `seq(to = 1, from = 3)`

Q&A 的時間又到囉

Q 成千上萬的函數哪學得完？

A 不用學完！沒人學得完！學常用的就好。

Q 函數的使用手冊看不懂耶。

A 我也常看不懂。儘量看，多嘗試，特別是 Example 部份。

Q 如何找能做某件事的函數？

A 請 Google 大神幫你找最快。真的。

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理**
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

檔案下載

- ① 至 <http://goo.gl/YRC0aN> 下載以下三個檔案。
2015RforBeginnerSlide.pdf 本投影片
exam.xlsx 例範資料
text.xlsx 練習資料
- ② 在 C disk 下創建一個 LearnR2015 資料夾，並將三個檔案置入該資料夾中。

讀取 CSV 資料檔案

- 1 開啟 `exam.xlsx`，注意第一列必需有一列「變數名」。
- 2 另存新檔 → 檔名為「exam」，類型為「CSV」，儲存在相同資料夾中。
- 3 `getwd()` 顯示目前 R 所在的路徑。
- 4 `setwd("C:/LearnR2015")` 到達該資料夾。
- 5 `dt <- read.csv("exam.csv")` 或是
`dt <- read.csv("C:/LearnR2015/exam.csv")` 以讀取該檔成為一個資料框 (data frame)，並取名為 `dt`。
- 6 `dt` 的結果是什麼？

提取特定變數（欄）

```
> dt
  ID Gender Group Literature Science
1 23      m     A         36      63
...
```

如何取得 Science 變數？直接輸入 Science 是不行的，因為它是在 dt 裡的變數。

- `dt$Science` 意思是「dt 裡的 Science 變數」
- `dt[, 5]` 意思是「dt 裡的第 5 欄變數」
- `attach(dt)` 可使 dt 的所有變數傳至表層。

提取特定重覆數（列）

- `dt[3 ,]` 取得 dt 裡的第 3 列變數
- `dt[c(3, 6) ,]` 取得 dt 裡的第 3 及第 6 列變數
- `subset(dt, Gender == "m")`
取得 Gender 是 m 的資料。
- `subset(dt, Science >= 60)`
取得 Science 大於等於 60 的資料。
- `subset(dt, Gender == "m" | Science >= 60)`
取得 Gender 是 m 或 Science 大於等於 60 的資料。
- `subset(dt, Gender == "m" & Science >= 60)`
取得 Gender 是 m 且 Science 大於等於 60 的資料。

Q&A 的時間又到囉

Q 可否直接讀取 `xlsx` 檔？

A 可以！請日後自行研究 `xlsx` 這個套件。

Q 可不可以資料排序？

A 可以！請日後自行研究 `order()` 和 `sort()`。

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖**
- 5 學習心得與討論資源
- 6 小試身手

描述性統計

常見的描述性統計函數

```
length(dt$Science)  #個數  
mean(dt$Science)    #平均數  
sd(dt$Science)      #標準偏差  
quantile(dt$Science) #百分位數
```

如果要求各組的描述性統計呢？使用 `tapply()`。

`tapply()` 的基本語法

```
tapply(變數, 分組因子, 運算函數, ...)
```

例如，要計算 Science 在不同 Gender 內的平均數：

```
tapply(dt$Science, dt$Gender, mean)
```

單樣本 T 檢驗

目標：檢驗 Science 的平均是否為 60。

t.test() 的基本語法

```
t.test(資料, alternative = "t" 或 "l" 或 "g",  
       mu = 假說平均數, ...)
```

#雙尾：

```
t.test(dt$Science, alternative = "t", mu = 60)
```

#右單尾：

```
t.test(dt$Science, alternative = "g", mu = 60)
```

#左單尾：

```
t.test(dt$Science, alternative = "l", mu = 60)
```


成對樣本 T 檢驗

目標：檢驗 Literature 和 Science 差之平均是否為 0。

`t.test()` 的基本語法

```
t.test(資料1, 資料2,  
       alternative = "t" 或 "l" 或 "g",  
       mu = 假說中配對差的平均數, pair = T, ...)
```

預設雙尾；預設平均差為零

```
t.test(dt$Literature, dt$Science, pair = T)
```

獨立雙樣本 T 檢驗

目標：檢驗二種 Gender 的 Literature 之平均是否相等。

t.test() 的基本語法

```
t.test(資料1, 資料2, alternative = "t" 或 "l" 或 "g",  
       mu = 假說中平均數的差, var.equal = T 或 F, ...)
```

```
t.test(應變數 ~ 僅二類的類別因子, data = 資料框, ...)
```

```
t.test(subset(dt, Gender == "m")$Literature,  
       subset(dt, Gender == "f")$Literature,  
       var.equal = T)
```

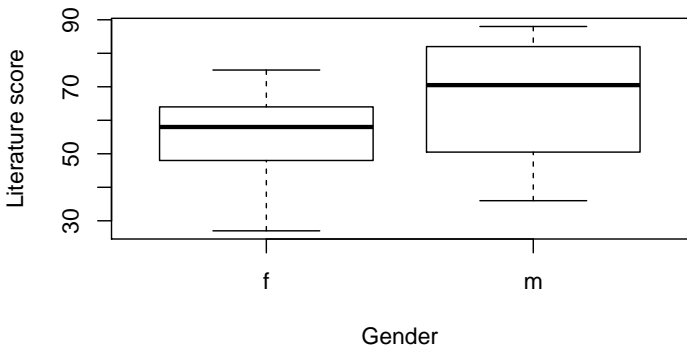
```
t.test(Literature ~ Gender, data = dt,  
       var.equal = T)
```

盒形圖

boxplot() 的基本語法

```
boxplot(應變數 ~ 類別因子, data = 資料框, ...)
```

```
boxplot(Literature ~ Gender, data = dt,  
        ylab = "Literature score", xlab = "Gender")
```



單因子變異數分析

目標：檢驗三種 Group 的 Literature 之平均是否相等，並進行 Tukey 事後檢驗。

aov() 和 TukeyHSD() 的基本語法

```
aov( 應變數 ~ 三組以上類別自變數, data = 資料框, ... )  
TukeyHSD(aov物件, "分組因子", ...)
```

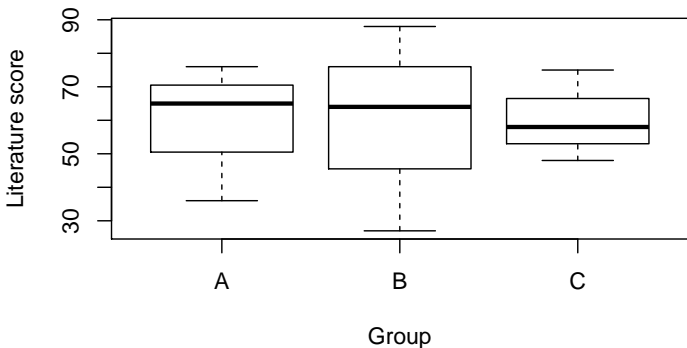
```
fit.1 <- aov(Literature ~ Group, data = dt)  
summary(fit.1) # Type I sum of square  
TukeyHSD(fit, "Group")
```

盒形圖

boxplot() 的基本語法

```
boxplot(應變數 ~ 類別因子, data = 資料框, ...)
```

```
boxplot(Literature ~ Group, data = dt,  
        ylab = "Literature score", xlab = "Group")
```



簡單線性迴歸

目標：建立 Science 對應 Literature 的簡單線性迴歸模型，並檢驗斜率是否為零。

lm() 的基本語法

lm(應變數 ~ 連續自變數, data = 資料框, ...)

```
fit.2 <- lm(Literature ~ Science, data = dt)
summary(fit.2)
anova(fit.2) # Type I sum of square
```

簡單線性相關

目標：計算 Science 與 Literature 的簡單線性相關係數是否為零。

`cor.test()` 的基本語法

```
cor.test(第一群資料, 第二群資料,  
         alternative = "t" 或 "l" 或 "g", ...)  
cor.test(~ 第一群資料 + 第二群資料, data = 資料框, ...)
```

```
cor.test(dt$Literature, dt$Science)
```

```
cor.test(~ Literature + Science, data = dt)
```

```
cor.test(~ Science + Literature, data = dt)
```

散佈圖

coef() 的基本語法

```
coef(lm物件, ...) # 取出各迴歸係數
```

plot.formula() 和 abline() 的基本語法

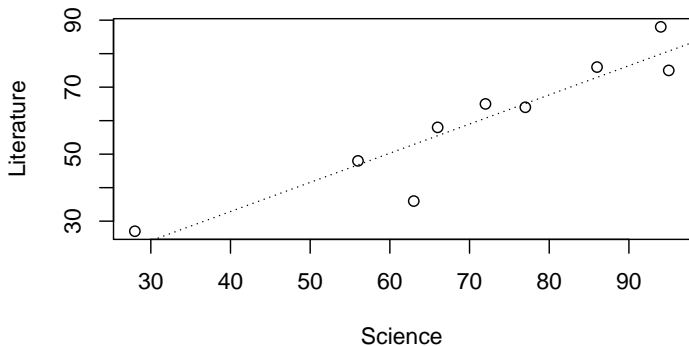
```
plot(縱軸資料 ~ 橫軸資料, data = 資料框, ...)6  
abline(a = coef(迴歸物件)[1], b = coef(迴歸物件)[2],  
       lty, col, ...) # 畫上迴歸線
```

```
plot(Literature ~ Science, data = dt)  
abline(a = coef(fit.2)[1], b = coef(fit.2)[2],  
       lty = 3)
```

⁶ plot.formula() 可簡寫成 plot()。

散佈圖

```
plot(Literature ~ Science, data = dt)  
abline(a = coef(fit.2)[1], b = coef(fit.2)[2],  
       lty = 3)
```



Q&A 的時間又到囉

Q 我怎麼知道我做對了？

A 拿出你的統計學課本的例題，用 R 做做看。

Q R 畫的圖想做更多調整……

A1 這件工作不是非常容易，需要經驗。有空看看 `par()` 和 `plot()` 的使用手冊。

A2 初學者可以先用 R 畫好概要圖，再以其它圖片編輯軟體後製。

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

阿盤的個人學習心得

- 修習使用 R 的課。
- 多「玩」。把函數裡的 Example 玩一玩、改一改。
- 肯問人。逛逛網路教學和論壇。
- 買（可能不只一本）書。
- 拿出統計學課本的例題，用 R 做做看。
- 做過的程式碼要建檔，方便日後使用。
- 卡關時，先用英文問 Google 大神。
- 做出答案時，不要直接相信這是正解，應該以專業人士、書籍、網頁資料驗證。

中文書籍推薦

繁體中文書非常少，但簡體中文書不少。去圖書館或書局翻翻。能看懂有收穫就有參考價值。初學程式語言者應該都需要一本。

- 《R 軟體：應用統計方法》陳景祥著，東華出版社。
對初學者很有幫助的一本。R 語言和統計學併重。
- 《R 錦囊妙計》Paul Teetor 著，張夏菁譯，歐萊禮出版社。
前半本內容是 R 語言，後半本是以 R 進行統計工作。
- 《R 语言实用教程》薛毅、陈立萍著，清华大学出版社。
- 《统计建模与 R 软件》薛毅、陈立萍著，清华大学出版社。
以數理統計為主，R 語言實作為輔。

英文書籍推薦

英文書選擇極多。我推薦以下幾本我喜歡或值得閱讀的。

- “Biostatistical Design and Analysis Using R: A Practical Guide” by Murray Logan. Wiley-Blackwell Press.
實驗設計和 R 並重，非常推薦。
- “The R Book, 2nd Edition” by Michael J. Crawley. Wiley Press.
較不易閱讀，但仍值得細讀。R 語言和統計併重。
- “A First Course in Statistical Programming with R” by W. John Braun & Duncan J. Murdoch. Cambridge University Press.
易讀。統計學基礎內容為主，但實驗設計部份少。

網路教學

- 《R 演習室》 @ youtube.com⁷
針對初學者的 R 視訊教學系列。有廣告（歡迎多看幾秒），但有提供影片載點。
- “Quick-R” by Robert I. Kabacoff⁸
- 請自行 Google 「R tutorial」，實在太多了。
- 中文教學網頁極少。還是找英文的比較實在。

⁷ <https://www.youtube.com/playlist?list=PL5AC0ADBF65924EAD>

⁸ <http://www.statmethods.net/>

網路討論區

- PTT 的 R_Language 板
路徑：戰略高手 → CompScience → R_Language
有幾位高手。對初學者友善。歡迎來和安德魯大大認親。
- R 軟體使用者論壇⁹
- Tag “R” @ stackoverflow.com¹⁰

⁹ <https://groups.google.com/forum/?hl=zh-TW#!forum/taiwanruser>

¹⁰ <http://stackoverflow.com/questions/tagged/r>

R 的套件

什麼是套件 (package) ?

安裝在 R 系統裡的外掛，讓你「不用重新造輪子」。

如何安裝、更新及引入套件？

- 連上網路之後，輸入 `install.packages("套件名稱")` 可以安裝某套件
- 在已安裝某套件之後，輸入 `library(套件名稱)` 可引入該套件，之後才可以使用它的功能。
- 連上網路之後，輸入 `update.packages()` 可以更新所有已安裝套件。

R 的官方套件庫

R 官方收錄有六千多個的套件，¹¹ 可直接以 `install.packages()` 安裝。

我常用的套件

- (一般／廣義) 線性模型：gmodels、lme4、aod
- 混合模型：lme4、nlme、MCMCglmm
- 蒙地卡羅、隨機化：permute、boot
- 多變量、群落生態、生物多樣性：vegan

¹¹ http://cran.r-project.org/web/packages/available_packages_by_name.html

Q&A 的時間又到囉

Q 如何找能做某件事的套件？

A 請 Google 大神幫你找最快。真的。

Q 阿盤學多久才叫「上手」、「有生產力」？

A 自學半年以上，但我今天就要把八成功力都傳給你了！

Q 聽到這裡，我想認輸了…… 我想重回用滑鼠搞定的世界。

A 只要是適合自己的工具，就是好工具。

大綱

- 1 R 簡介與操作環境
- 2 R 的函數
- 3 資料的讀取與整理
- 4 統計分析與繪圖
- 5 學習心得與討論資源
- 6 小試身手

今日的總複習

- 建立一個（適合自己的）R 工作環境
- 了解 R 的函數與如何閱讀其使用手冊
- R 如何讀取並整理資料
- 練習常見的統計方法
- 讓自己更厲害的資源

按今日課程試著完成以下練習

- 1 想辦法以 R 讀取 `nation-data.xlsx` 的內容並命名為 `mydt0` 資料框。以檔案中所有國家為樣本完成以下分析。
- 2 利用配對樣本 T 檢驗，考驗 `Mortality.rate.child` 之平均是否顯著高於 `Mortality.rate.newborn` 之平均。提示：不是雙尾檢驗。
- 3 以 `GDP.10000` 為組別，計算 `HIV.rate` 在各組的平均值和標準偏差，並利用獨立雙樣本 T 檢驗比較組間的平均是否顯著不等，以及繪製對應的盒形圖。
- 4 以 `Continent` 為組別，計算 `Age.ave` 在各組的平均值和標準偏差，並利用單因子變異數分析比較組間的平均差異是否顯著不等，以及繪製對應的盒形圖。
- 5 以 `HIV.rate` 為反應變數（應變數），`Age.ave` 為解釋變數（自變數），建立簡單線性迴歸模型，並檢驗斜率及相關係數是否顯著不為零，以及繪製對應之散佈圖。

參考解法 I

讀入 CSV 檔：

```
> setwd("somewhere") # 更變目前路徑
> mydt0 <- read.csv("nation-data.csv") # 讀檔
> mydt0
```

	Nation	Continent	HIV.rate	Age.ave	Mortality.rate.child	...
1	Algeria	1Africa	0.10	72.904	18.090586	...
2	Morocco	1Africa	0.10	71.882	13.156450	...
3	Zambia	1Africa	13.50	48.513	105.545396	...

71	Slovak Republic	4Europe	0.06	75.242	4.203745	...
72	Latvia	4Europe	0.70	73.039	4.064162	...

```
> names(mydt0) # 查看變數名
```

[1] "Nation"	"Continent"	"HIV.rate"
[4] "Age.ave"	"Mortality.rate.child"	"Mortality.rate.newborn"
[7] "GDP.10000"		

```
> dim(mydt0) # 查看列數與欄數
```

[1] 72 7

參考解法 II

對 Mortality.rate.child 和 Mortality.rate.newborn 的配對樣本 T 檢驗：

```
> x.1 <- mydt0$Mortality.rate.child  
> x.2 <- mydt0$Mortality.rate.newborn  
> t.test(x.1, x.2, paired = T, alternative = "g")  
> t.test(x.2, x.1, paired = T, alternative = "l") # 也可
```

Paired t-test

data: x.1 and x.2

t = 2.1011, df = 71, p-value = 0.01959

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

0.8812246 Inf

sample estimates:

mean of the differences

4.260981

參考解法 III

以 GDP.10000 分組對 HIV.rate 之描述：

```
> tapply(mydt0$HIV.rate, mydt0$GDP.10000, mean))  
> with(mydt0, {tapply(HIV.rate, GDP.10000, mean)}) # 亦可
```

high	low
0.286087	1.213061

```
> with(mydt0, {tapply(HIV.rate, GDP.10000, sd)} )
```

high	low
0.3095707	2.7004554

參考解法 IV

以 GDP.10000 分組對 HIV.rate 之獨立雙樣本 T 檢驗：

```
> t.test(HIV.rate ~ GDP.10000,  
         data = mydt0, var.equal = T)
```

Two Sample t-test

data: HIV.rate by GDP.10000

t = -1.6351, df = 70, p-value = 0.1065

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.0576478 0.2036993

sample estimates:

mean in group high mean in group low

0.286087

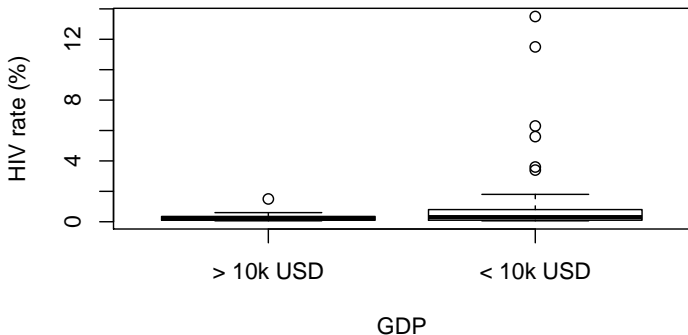
1.213061

不過此例 `t.test(..., var.equal = F)` 可能較洽當，因為二組的變方差距不小。

參考解法 V

以 GDP.10000 分組對 HIV.rate 之盒形圖：

```
> boxplot(HIV.rate ~ GDP.10000, data = mydt0,  
+         xlab = "GDP", ylab = "HIV rate (%)",  
+         xaxt = "n")  
> axis(1, 1:2, label = c("> 10k USD", "< 10k USD"))
```



參考解法 VI

以 Continent 分組對 Age.ave 之描述：

```
> with(mydt0, {tapply(Age.ave, Continent, mean)})
```

```
1Africa 2America    3Asia 4Europe  
61.11923 74.48475 72.31782 77.37283
```

```
> with(mydt0, {tapply(Age.ave, Continent, sd)})
```

```
1Africa 2America    3Asia 4Europe  
9.308895 4.014003 6.383229 3.820449
```

參考解法 VII

以 Continent 分組對 Age.ave 進行單因子變異數分析：

```
> f.anova <- aov(Age.ave ~ Continent, data = mydt0)
> summary(f.anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Continent	3	2439	813.0	24.12	9.66e-11 ***
Residuals	68	2292	33.7		

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

註：此例之間變方甚不同質，故以 `oneway.test(Age.ave ~ Continent, data = mydt0)` 進行組間變方不同質之修正，或是以 `kruskal.test(Age.ave ~ Continent, data = mydt0)` 進行 Kruskal-Wallis rank sum test，可能較洽當。

參考解法 VIII

Tukey 事後檢驗：

> TukeyHSD(f.anova, "Continent")

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = Age.ave ~ Continent, data = mydt0)
```

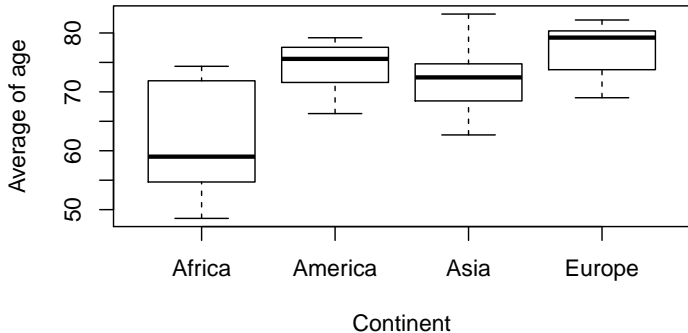
```
$Continent
```

	diff	lwr	upr	p adj
2America-1Africa	13.365519	7.2440030	19.487035	0.0000014
3Asia-1Africa	11.198593	5.5646064	16.832579	0.0000102
4Europe-1Africa	16.253603	11.1760641	21.331141	0.0000000
3Asia-2America	-2.166926	-7.9324031	3.598550	0.7556740
4Europe-2America	2.888083	-2.3349728	8.111139	0.4693185
4Europe-3Asia	5.055010	0.4129029	9.697117	0.0275116

參考解法 IX

以 Continent 分組對 Age.ave 繪製盒形圖：

```
> boxplot(Age.ave ~ Continent, data = mydt0,  
+         xlab = "Continent", ylab = "Average of age",  
+         xaxt = "n")  
> axis(1, 1:4,  
+      label = c("Africa", "America", "Asia", "Europe"))
```



參考解法 X

HIV.rate vs Age.ave 的簡單線性迴歸：

```
> fit.reg <- lm(HIV.rate ~ Age.ave, data = mydt0)
> summary(fit.reg)
```

Call:

```
lm(formula = HIV.rate ~ Age.ave, data = mydt0)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.6995	-0.8609	-0.0631	0.7118	7.8572

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.09700	1.73463	8.703	9.27e-13 ***
Age.ave	-0.19488	0.02369	-8.225	7.03e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.63 on 70 degrees of freedom

Multiple R-squared: 0.4915, Adjusted R-squared: 0.4842

F-statistic: 67.65 on 1 and 70 DF, p-value: 7.027e-12

參考解法 XI

HIV.rate vs Age.ave 的簡單線性相關：

```
> cor.test( ~ HIV.rate + Age.ave, data = mydt0)
> cor.test(mydt0$HIV.rate, mydt0$Age.ave) # 亦可
```

Pearson's product-moment correlation

data: mydt0\$HIV.rate and mydt0\$Age.ave

t = -8.2253, df = 70, p-value = 7.027e-12

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.8024053 -0.5604066

sample estimates:

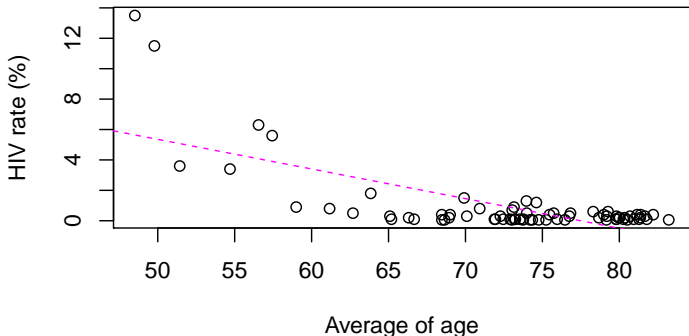
cor

-0.7010578

參考解法 XII

HIV.rate vs Age.ave 的散佈圖：

```
> plot(HIV.rate ~ Age.ave, data = mydt0,  
+       xlab = "Average of age", ylab = "HIV rate (%)")  
> abline(a = coef(fit.reg)[1], b = coef(fit.reg)[2],  
+        lty = 2, col = 6)
```



謝謝收看

```
> cat("See you next time :) \n")
```

```
See you next time :)
```

```
> q()
```

```
Save workspace image? [y/n/c]: n
```