

110年衛生福利資料科學中心


R軟體推廣課程 《進階篇》

課程前導

講師：劉品崧 統計分析師

花蓮慈濟醫院

版權宣告

- 本次課程的所有內容依據台灣創用CC條款進行授權
 - 講義
 - 程式碼
 - 範例資料，但不包含衛生福利部統計處提供之模擬資料檔
- 授權類別 
 - 姓名標示—非商業性—禁止改作
 - 本授權條款允許使用者重製、散布、傳輸著作，但不得為商業目的之使用，亦不得修改該著作。使用時必須按照著作人指定的方式表彰其姓名。

課程地圖

- **使用GGPLOT2繪圖** 2021-05-07(五) **慈濟大學** C305電腦教室
 - 套件簡介、統計圖形、客製化圖形外觀
- **存活分析** 2021-05-10(一) **中山醫學大學** 正心樓6樓0622電腦教室
 - 模型介紹、圖形與模型分析、衍伸模型
- **資料的群組操作** 2021-05-31(一) **國立成功大學** 電算中心75309電腦教室
 - 清單(list)物件、迴圈撰寫、apply與function
- **建立預測模型** 2021-06-28(一) **國家衛生研究院** 圖資大樓5樓L5203電腦教室
 - 因子選擇、建立模型、驗證模型
- **共病指標的資料整理** 2021-06-30(五) **台北醫學大學** 杏春樓2樓A電腦教室
 - CCI介紹、資料整理程式撰寫、統計分析

課程進行模式

- 109年基礎篇

- 講義說明 + 學員動手執行範例程式碼
- 學員必須能自行完成基礎操作，或趕緊複習109年基礎教材
 - 安裝R、Rtools、Rstudio、下載套件、載入套件
 - R語言指令撰寫、匯入資料、常用套件 e.g. `fst`、`data.table`

- 110年進階篇

- 以**真實研究資料**作為**背景情境**
- 以**研究過程實務**作為**教學目標**

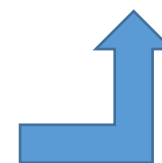
課程錄影



109年度R基礎課程-劉品崧老師



衛生福利資料科學中心慈濟大學研究分中心



講義教材







<https://github.com/PSLiu/NHIRD-R-2020>

<https://github.com/PSLiu/NHIRD-R-2021>

R及相關軟體的下載及安裝

- 依據所使用的作業系統不同，你需要安裝以下軟體
- R系列皆為免費開源軟體，Google搜尋即可找到下載點

					
安裝軟體		R	Rtools	Xcode	RStudio
作業系統	Windows (務必以系統管理員身分執行安裝)	V	V		V
	Mac	V		V	V
	Linux	V			V

課前準備事項

- 自行安裝R、RStudio與相關軟體
 - RStudio更改Encoding為UTF-8
- 下載完整教材檔案並解壓縮
 - 讓資料夾路徑越單純越好，不要有中文或是放在OneDrive裡面
- 修改資料夾路徑物件為自己的

課程進行方式 for 雲端授課模式

- 時間分配概略比例
 - 老師講課30分鐘，學員練習10分鐘，老師解說與學員Q&A再10分鐘
 - 下課休息10分鐘
- 線上授課配合學員自行操作
 - 請各位在電腦上使用R跟著老師run範例程式 & 動手練習
 - 如果遇到Bug或是Error，請準備開麥克風提問及分享畫面
讓老師遠端協助解答你的問題

課程提問模式 for 雲端授課模式

- 有問題，隨時可以開麥克風提問！
- 提問小技巧：請簡述提問者 & 問題描述(位置/狀況)
 - 公衛系張老師，講義第21頁
 - 神經外科林醫師，程式碼的OO段落
 - 醫學系張同學，執行產生ERROR
- 如果是比較前面的進度想要再聽一次，請在討論區留言
 - 提問者請在線上稍後，我們會找時間回答！

課程演練資料情境介紹：肺炎住院及死亡預測

- 使用資料集：**pneumonia.fst**

- 2014年6月的住院申報檔案，任一診斷欄位、具有肺炎(pneumonia)的 ICD-9-CM診斷：480.xx ~ 486.xx的住院病人，觀察其存活狀況

- 資料結構

- 性別、年齡、投保、住院資訊、死亡日期、病史、CCI

- Study outcome

- 住院天數(continuous)、住院天數是否超過7/14/21天(binary)
 - 出院後半年內是否死亡(binary)、出院後存活分析(time-to-event)

課程演練資料情境介紹：糖尿病病人MPR預測

- 使用資料集：**dmmpr.fst**
 - 2014年6月的門診、藥局申報資料中有使用糖尿病藥物(ATC Code屬於A10開頭的藥物)的病人，觀察其下半年(7-12月)的藥物遵從性(MPR)
- 資料結構
 - 性別、年齡、投保、MPR、過去病史、CCI
- Study outcome
 - 糖尿病藥物遵從性(continuous)
 - 糖尿病藥物遵從性(binary, 1 if $MPR \geq 80$)

獨立作業區R軟體與相關應用條件

- R軟體版本
 - 目前最新為**4.1.0**，亦有舊版，最早為**3.6.3**
 - 僅包含R預設的30個基本套件
- RStudio軟體
 - 統計處會統一更新至最新版本，目前為**1.3.1093**
- 目前電腦硬體條件(統計處會視情況調整)
 - 純R的桌面配備**40 GB**的記憶體
 - STATA+R為**40 GB**，SAS+R為**16 GB**

VDI預約管理網站

使用軟體

請選擇

請選擇

R

SAS (含IML) +R

SAS+R

SAS+SPSS+R

SAS+STATA+R (STATA版本為15版)

SPSS+R

STATA+R (STATA版本為14版)

Transfer+R

如何在獨立作業區更有效地使用R軟體？ 1/3

- 填寫SEC003表單，攜入使用者所需套件以及Rtools軟體

SEC003

中華民國 104 年 8 月 28 日修訂

序號：_____

衛生福利資料科學中心使用軟體聲明書

計畫主持人名稱

計畫案號
及完整名稱

本人 000 因執行 H109000 XXXXXXXXXXXXX 研究計畫所需，擬於衛生福利資料科學中心(以下簡稱貴中心)電腦安裝下述版本之軟體：R 軟體(v. 4. 0. 3)之相關套件共 428 件(詳如附件)，以及 RTools 軟體(v. 40)，以利本計畫之執行。

本人聲明前述軟體確係合法使用，若有侵犯任何第三人之智慧財產權，其法律責任概由本人負責；且前述軟體之安裝，若有造成貴中心電腦或系統有任何損害，願付損害賠償責任。

攜入套件數量
對應R版本
Rtools軟體

如何在獨立作業區更有效地使用R軟體？ 2/3

- 整理完整套件清單
 - 建議在自己的電腦上先把所有已下載套件刪除，重新安裝適用且對應獨立作業區的R軟體版本，並執行已下指令下載所需套件，儲存為清單

```
# 將想要攜入的套件名稱變成物件
a <- c("fst", "data.table", "tidyverse")

# 下載
install.packages(a, dependencies = TRUE)

# 取得所有下載套件的名稱(包含相依套件)
b <- list.files(.libPaths()[1])

# 將所有清單寫出成為清單
write.csv(b, file = "PackagesList_basic.csv")
```

如何在獨立作業區更有效地使用R軟體？ 3/3

- 安裝套件之前要先安裝Rtools，再打開RStudio，執行以下指令

```
# 將想要安裝的套件名稱變成物件  
a <- list.files("I:/3 IN/H109268/YourFolder", pattern = ".zip$",  
full.names = TRUE)  
  
# 使用迴圈逐一安裝  
install.packages(a, repos = NULL, type = "win.binary")
```