

國立中山大學政治學研究所
經驗研究與資料分析
Empirical Studies and Data Analysis
(IPS505)

Spring 2015 (1032)

授課教師：劉正山 副教授	Instructor: Prof. Frank Cheng-shan Liu
上課時間：星期四 下午 2:10-5:00	課程討論時間：星期一、五 下午1:00-3:00 或另約
上課教室：社 3008	電子郵件信箱： csliau@mail.nsysu.edu.tw
教師研究室：社 1009-1	校內分機：5555
教師個人網站： http://goo.gl/DRZjXA	
**重要：Evernote筆記本 (閱覽及下載課程相關教材檔案)： http://goo.gl/VpVioS	

課程簡介

社會科學發展很重要的基礎是對經驗現象的描述和解釋，而1960年代以來發展成熟的假設檢證途徑可說是目前所有基礎科學、跨學門所共享的求知方法。這個方法結合二十多年來電腦和資訊科學的發展，已經使得操作軟體進行資料分析，並由此發掘（描述）問題、提出解釋，成為發展理論學者的必備能力。本課程教導同學使用R-studio軟體結合R語言進行社會科學的資料分析。學會這些方法的同學將能夠正確解讀以資料分析為主要研究方法的論文、重製（replicate）重研究發現，以及進行自己有興趣的分析工作，進一步分析其他學科及未來工作領域的資料。

本課程為本所必修課碩士班「政治學研究方法」的延伸，與本所選修課「政治學方法論」、「民意與選舉行爲研究」、「政治傳播」等課程互補。本課程著重於訓練同學操作軟體的能力，並輔以深入淺出的統計原理介紹。全學期課程內容包括以下幾個主題：一、認識資料與資料型態；二、認識資料分析的工具；三、概念的測量與編碼；四、模型的建立與假設檢證；五、分析結果的呈現與詮釋。

課程教材與指定閱讀

- 指定專書/專章（放置於課程筆記內）：
 - Zumel, N., & Mount, J. (2014). Practical Data Science with R (1 edition.). Manning Publications. [Chapters 1-5, & 7]
 - Kabacoff, R. (2011). R in Action. Shelter Island, NY: Manning Publications. [Chapters 1-8 & 13]

- 本課程教材內容亦參考以下專書：
 - Dalgaard, P. (2008). *Introductory Statistics with R*. (2nd Ed.). New York, NY: Springer-Verlag.(中山大學電子書)
 - Gerbing, D. W. (2013). *R Data Analysis without Programming* (1 edition). New York: Routledge.
 - Lander, J. P. (2014). *R for Everyone: Advanced Analytics and Graphics*.
 - Teetor, P. (2011). *R Cookbook* (1st ed.). O'Reilly Media.
 - Verzani, J. (2005). *Using R for Introductory Statistics*. Chapman & Hall/CRC.
 - * 免費版下載：Verzani, J. (2002). *SimpleR: Using R for Introductory Statistics*. <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>.
 - Verzani, J. (2011). *Getting Started with RStudio*. O'Reilly Media.

關於統計分析語言R

近年來，統計用的程式語言R受到跨領域學者的支持，已成為全球投入資料科學（data science）發展的共同語言和工具，而此趨勢在國內亦已受到產、官、學界的重視。R是由國外一群熱心於統計與資訊的專業學者、志工們所共同努力發展的一套免費科學計算語言，學者使用它創造出了數千個針對不同學科處理資料問題時需要的套件軟體，並以開放源始碼的精神免費開放供全球使用。由於R及它的套件包含了絕大多數的統計運算功能，且適用於各種作業平台(MS Windows, Mac OS X, Linux)，使它愈來愈具有取代費用高昂的商用軟體（如SAS, SPSS）的潛力。如此，R已成為Big Data產業公認、適合作為雲端運算的企業營運工具，甚至熟用R已儼然成為美國頂尖大學科系的畢業條件或是必備的能力。

這門課是從「資料」的角度而非「統計」的角度所設計的入門課程，因此修課同學不需要任何統計學背景，而課堂中的統計知識，也是以最淺白的方式讓同學達到可以正確判讀及詮釋分析結果的水準。即使如此，老師鼓勵同學透過修校內外及所外其他課程的方式，補強統計學的學理。若對此領域有興趣的同學，可在本學期之後繼續修習本課程的進階課程「政治學量化資料分析」。

評分標準

- 三次作業（20%, 20%, 20%）：依課程進度繳交課後練習作業。第四次作業為選擇性作業，可選擇之後擴充為學期報告。
 - 第一次作業：錨定資料與問題意識。請從國內外資料庫中（如中研院學術調查研究資料庫）挑選一個你有興趣的資料，在作業中報告：
 - * 簡介這個資料檔
 - * 指出資料檔中那個（或那幾個）變數是你覺得有興趣要進行分析研究的「依變數」，那些是你想挑選作為解釋的「自變數」
 - * 列出至少三篇與你挑選的依變數有關的文獻
 - * 建議使用markdown書寫。

- 第二次作業：變數的編碼與模型的建立。請建立一個解釋模型並將資料檔的讀入匯出、變數的編碼及無效值的清理等步驟寫入程式語法檔（檔名為Variables.R）中。並在同一個語法檔內，說明：
 - * 自變數（至少要有兩個）與依變數之間的關係（為什麼你會認為該自變數會影響依變數）
 - * 控制變數與依變數的邏輯關係（為什麼該控制變數會影響依變數）。
 - * 必須加上「檔頭」說明，並使用#來作註記
- 第三次作業：變數之間關聯檢定。請依第二次作業所提供的預期，
 - * 寫下三個或三個以上、六個以下的假設
 - * 使用卡方檢定和交叉列聯表，檢定它們之間的關係
 - * 依各自的假設寫下分析的結果報告
 - * 請將資料檔的讀入、匯出、變數的編碼及無效值的清理等步驟寫入程式語法檔（R檔）並命名為Variables.R。將你的分析，如使用load()及CrossTable()等指令寫在另一個語法檔內，命名為Analysis.R。
 - * 將這兩個R檔以Email附檔的方式寄給老師。
 - * 必須善用註記符號來呈現必要的資訊（如檔頭的姓名及編碼的說明）。
- 第四次作業：假設檢證與分析結果詮釋。請依據第三次作業製作Analysis.R與Variables.R的經驗，製作幾個二元勝算對數模型，並比較模型之間的差異。
 - * 依變數必須是二元變數，且每個模型的自變數與控制變數不得少於三個。
 - * 必要要有Variables.R及Analysis.R兩個檔案
 - * 另加上一個結果報告檔，以表格呈現分析結果，並加上適當的分析及討論（繳交格式可為pdf, Rmd, html, 或md）。
- 學期報告（20%）：以重製論文發現或使用自選資料檔進行探索性研究的學期作品（繳交格式可為pdf, Rmd, html, 或md）。
- 課程參與（20%）：課程出席、準備、互動與上網使用教材的綜合表現。

課程進度

1. [2/26] 課程介紹

- 談談經驗研究與資料分析的目的
 - 政府開放資料的趨勢
 - 學門內部的資料分析趨勢：社會變遷、民主化、傳播調查資料庫
- Check this out: Coursera for Data Analysis (<https://class.coursera.org/dataanalysis-001/lecture/index>)
- Recommend the use of cloud tools such as Evernote and Dropbox (<http://db.tt/0MW1fE4>).
- 課後：安裝Tableau Public, R, and R-Studio

2. [3/5] 認識資料與資料型態

- 認識資料與資料型態
- 資料庫的使用

3. [3/12] 資料的描述與製圖

- R的基本操作 & R studio的基本操作
- Tableau Public與Rstudio的操作
- 資料的視覺化與說故事（依資料描述經驗現象或點出問題）

4. [3/19] 老師出席ISGC2015國際會議-停課乙次

- 提前至3/17補課：出席Prof. Francis Lee演講
 - “Digital media and social movements: Self-mobilization and its limitations”
 - 3/17, 14:00-16:00
 - 地點將另行通知

5. [3/26] 基本觀念簡介/複習

- 工具：markdown editors
- 基本觀念：
 - 資料框（data frame）、分析單元（unit of analysis）、樣本（sample）、母體（population）、抽樣（sampling）、模型（model）、變數（variables）、測量（measurement）
 - 理論、模型、假設之間的關係
 - 交叉變數、控制變數與與中介變數：
 - * Kam, C. D. & Franzese, R. J. (2007). Modeling and Interpreting Interactive Hypotheses in Regression Analysis. Ann Arbor: University of Michigan Press. [Chapter3]
- 模型的設計與修正
- 變數的編碼細節

6. [4/2] 老師出席WPSA2015國際會議-停課乙次

- 4/23正常上課作為補課
- 第一次作業繳交

7. [4/9] 概念的測量與編碼

8. [4/16] 變數相關性檢驗

- 卡方檢定

9. [4/23] 模型的建立與線性迴歸

- 期中考週原不上課，本週作為4/2補課；
- 第二次作業繳交

10. [4/30] 機率模型的建立與迴歸

- 二元勝算對數模型 (logistic regression analysis)

11. [5/7] 模型的校準及分析結果詮釋

12. [5/14] 表格製作與簡報

- 第三次作業繳交

13. [5/21] 實例研讀與重製

14. [5/28] 延伸主題 (1) : Factor/Item Analysis

15. [6/4] 延伸主題 (2) : dplyr與資料的管理

- 第四次作業繳交

16. [6/11] 學生研究成果分享 & 作業問題討論

17. [6/18] 課程總結

18. [6/23] 期末考週 & 期末報告截止日

其他補充/自修教材

- 統計入門讀物
 - David S. Moore (1998). Statistics (統計, 讓數字說話) 鄭惟厚譯。台北: 天下文化。
 - 鄭惟厚 (2007)。你不能不懂的統計常識。台北: 天下文化。
 - Lewis-Beck, M.S. (1995). Data Analysis: An Introduction. Thousand Oaks, CA: Sage Publications.
- 統計學推薦教材
 - [LC] 林惠玲、陳正倉。2008。基礎統計學 (二版)。台北: 雙葉書廊
 - 王濟川、郭志剛。2005。Logistic迴歸模型: 方法及應用。台北: 五南。
 - 林惠玲、陳正倉。2004。統計學: 方法與應用三版 (上) (下)。台北: 雙葉書廊。
- R入門的相關教材
 - Gerbing, D. W. (2013). R Data Analysis without Programming (1 edition.). New York: Routledge.

- Lander, J. P. (2013). R for Everyone: Advanced Analytics and Graphics (1 edition.). Upper Saddle River, New Jersey: Addison-Wesley Professional.
 - Muenchen, R. A. R for SAS and SPSS Users New York, NY : Springer-Verlag, 2008.(中山大學電子書)
 - Venables, W.N., Smith, D.M., & the R Development Core Team (2005). An Introduction to R. Free download: <http://cran.r-project.org/doc/manuals/R-intro.pdf>.
- 進階參考書
- Fox, J. (2002). An R and S-Plus companion to applied regression. Sage Publications.
 - Gelman, A. & Hill, J. Data Analysis Using Regression and Multilevel/Hierarchical Models. UK : Cambridge University Press, 2006.(中山大學電子書)
 - Harrell, F. E. (2001). Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. Springer.
 - Horton, N. J., & Kleinman, K. (2010). Using R for Data Management, Statistical Analysis, and Graphics (1st ed.). CRC Press.(中山大學電子書)
 - Janert, P. K. (2010). Data Analysis with Open Source Tools (1st ed.). O'Reilly Media. Teetor, P. (2011). R Cookbook (1st ed.). O'Reilly Media.
 - Klass, Gary M. 2008. Just Plain Data Analysis: Finding, Presenting, and Interpreting Social Science Data.Lanham, MD: Rowman & Littlefield. [Chapters 2 and 3].
 - Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Sage. (鄭旭智、張育哲、潘倩玉、林克明譯。2002。類別與受限依變項的迴歸統計模式。台北：弘智文化。)
 - Spector, P. (2008). Data Manipulation with R. New York, NY: Springer Science. (中山大學電子書)