

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 2013 (1012)

---

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： <a href="mailto:cslu@faculty.nsysu.edu.tw">cslu@faculty.nsysu.edu.tw</a>
教師研究室：社 1009-1	校內分機：5555
課程公告網站： <a href="http://www2.nsysu.edu.tw/politics/liu">http://www2.nsysu.edu.tw/politics/liu</a>	
課程網站：請登入中山網路大學 <a href="http://cu.nsysu.edu.tw/">http://cu.nsysu.edu.tw/</a> （課程錄影檔）	

---

## 課程簡介

近年社會科學發展很重要的趨勢之一，是運用統計學和計量經濟學到研究議題的發掘和理論的發展上。假設檢證的概念與工具則成了跨學門共享知識的基本語言，而動手分析資料已成為一個從事經驗研究者能否自行檢證並發展理論的基礎能力。本課程教導同學使用商用套裝軟體（中山大學授權版）SPSS以及開放源碼（Open-Source）程式語言R來做資料分析的工作，包括使用SPSS進行資料讀取，以及使用R/R-studio進行初級資料分析（卡方）和進階資料分析（迴歸）。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。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。本課程所使用的資料包括國內外社會學與政治學（比較政治、選舉研究、與國際關係等領域）的資料，如World Value Survey, CSES與TEDS等。全學期課程內容包括：一、認識資料與資料型態，二、認識資料分析的工具（SPSS與R），三、概念的測量與編碼，四、模型的建立與假設檢證，五、與理論的對話，六、國內外經驗研究成果的評估與重製，七、獨立進行研究與報告寫作。

這門課是資料分析的入門課程，不要求任何統計學背景。本課程為本所必修課「政治學研究方法」的延伸，與本所選修課「政治學方法論」、「民意與選舉行為研究」、「政治傳播」互補。本課程的進階為「政治學量化資料分析」。

## 課程教材與指定閱讀

- 本課程教材為老師講義，內容來自以下專書：

- Dalgaard, P. (2008). Introductory Statistics with R. (2nd Ed.). New York, NY: Springer-Verlag.(中山大學電子書)
- 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.
- 本課程以影音串流的方式保存上課內容，除了在課堂上教導同學如何進行政治與社會的資料分析，錄製的課程內容將有助於同學複習及完成作業。使用影音串流的目的是在於訓練同學操作軟體的能力，減少課堂來不及筆記以及回家自行摸索的困擾。

## 關於統計分析語言R

近年來，開放源碼的觀念在全球電腦軟體界蔚為一股風潮，此趨勢在國內亦備受產、官、學界的重視。主要原因是它不僅可以大幅降低軟體使用的成本，它強大的功能更可以提升產業與學術研究的競爭力。國內外已經有許多倡導使用自由軟體的協會及社群陸續成立，各類型研討會持續開辦中，而政府也正積極在政策上與教育上推出各種相關計畫，期望藉由這些計畫的執行，帶動我國自由軟體的發展，最終將開發的軟體應用於產業界，以帶來經濟上的最大效益。目前商業統計軟體的價格非常昂貴，學校所採購的套件數量有限，且只能安裝於學校的教師研究室或學生實驗室內使用。此外，由於作業系統上的差異，不同作業系統的電腦經常需要重複購買同一款軟體來使用。這除了在教學及研究上造成極大的不便，對於學校的教學成本也造成極大的浪費。因此，崇尚自由、非以營利為目標的大學校園開始面對開放源碼軟體的使用。

R軟體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軟體，它包含了絕大多數的統計運算功能，適用於各種作業平台(MS windows, Mac OS X, Linux)，不僅具有可擴充的統計分析模組，而且具備完整的物件導向程式設計能力、強大的繪圖輸出功能、結合資料庫的使用，以及預設雙倍精確度的特性，足以應用在各式各樣的科學分析與工商決策上。任何使用者皆可以自由下載、修改與擴充，以滿足自己所需要的計算樣貌。整體而言，R非常適合作為雲端時代下的企業運算工具。R軟體目前已經被許多國際知名公司廣泛使用，而且熟用R已儼然成為美國許多大學相關科系學生的畢業要件。雖然國內目前使用的情形剛剛起步，然而已經有一些較具前瞻性的企業在招募品保工程師時，特別指明所招人才需要具備撰寫R程式的能力。可以預期的是，R這個具備強大計算功能的利器在未來的一、二十年間將快速發展成為一主流計算工具，並足以抗衡於目前所有的商用統計軟體。

## 本學期可供作業之資料檔

- 國際資料庫
  - World Value Survey

- CSESuni
- worldtreatyindex.com
- IMPACT survey: <http://impactsurvey.org/>
- 國內資料庫
  - TEDS 2010C
  - 東亞民主研究計畫-亞洲民主動態調查
  - 中研院學術調查資料庫: <https://srda.sinica.edu.tw/>
- 國外資料庫
  - List of public databases from the Washington Post
    - \* <http://www.washingtonpost.com/wp-srv/metro/data/datapost.html>
    - \* <http://www.inside-r.org/howto/finding-data-internet>
- 其他老師提供的電訪調查資料檔

## 評分標準

- 三次作業 (10%, 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。
- 學期報告 (30%) : 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。
- 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

### 第一部分：重要概念介紹與回顧

#### 1. [2月21日] 開場暖身

- 談談經驗研究與資料分析的目的
- 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>).

#### 2. [2月28日] 國定假日

- 暖身閱讀：
  - Box, G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, 74 (365), 1-4.

### 3. [3月7日] 認識資料與資料型態

- 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
- 第一次作業 (五頁內)：資料的基本認識。請依中山網路大學本課程網站上所提供的資料檔，挑選一個你有興趣的資料，報告這個資料檔吸引你的原因或你使用它的理由，以及它有什麼特色。

### 4. [3月14日] 認識資料與資料型態

- 欄位與變數
- 變數與假設檢證

### 5. [3月21日] 認識資料分析的工具 I (SPSS & R)

- SPSS介紹與大型資料檔的使用
- 資料格式與整理
- R的基本操作 & R studio的基本操作
- R-Comander與R-studio的操作介紹
- 在R中建立、匯出、匯入和儲存資料檔
- 轉換SPSS格式資料檔到R的資料格式
- 第一次作業繳交

### 6. [3月28日] 認識資料分析的工具 II (SPSS & R)

### 7. [4月4日] 校際活動

- 暖身閱讀：
  - The report by UCLA Statistical Consulting Group <http://www.ats.ucla.edu/stat/TechnicalReports/>
  - Altman, M. & McDonald, M.P. (2001). Choosing reliable statistical software. PS: Political Science & Politics, 34 (3) , 681-687.

## 8. [4月11日] 概念的測量與編碼

- 以SPSS&R-commander來進行變數的重新編碼 (recoding)
- 變數的編碼細節
- 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 (R檔) 中。並在同一個語法檔內，說明 (1) 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)，以及 (2) 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)。將此檔命名為Variables.R，以Email附檔的方式寄給老師。注意，你必須加上「檔頭」說明，並使用#來作註記。

## 9. [4月18日] 概念的測量與編碼

### 10. [4月25日] 模型的建立與假設檢證

- 期中考週；繳交第二次作業
- 使用SPSS與R進行卡方檢定

### 11. [5月2日] 模型的建立與假設檢證

- 使用R製作二元勝算對數模型 (logistic regression analysis)
- 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 ]
- 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，寫下三個假設。接下來使用卡方檢定和交叉列聯表，檢定它們之間的關係，並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。再將這兩個R檔以Email附檔的方式寄給老師。注意：你必須善用註記符號來呈現必要的資訊 (如檔頭的姓名)。

### 12. [5月9日] 與理論的對話

- 「因果」關係：
  - Holland, P. W. (1986). Statistics and Causal Inference. Journal of the American Statistical Association, 81 (396), 945-960.

### 13. [5月16日] 期刊研究成果的解讀

- 本週因老師出國開會 (5/13-5/19)，可能調整本週課程時間或內容。
- 第三次作業繳交。

#### 14. [5月23日] 期刊研究成果的解讀

- 補充閱讀：
  - Pennell, B. E., Shapiro, M., Tessler, M., Weir, D. & Willis, R. (2002). How Americans responded: A study of public reactions to 9/11/01. *PS-political Science & Politics*. 35 (3): 511-516.
- 第四次作業：請依據第三次作業製作Analysis.R與Variables.R的經驗，以及這兩週課程的內容，製作三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，每個模型的自變數與控制變數不得少於三個和多於六個。除了Variables.R及Analysis.R之外，另加上一個結果報告檔（doc或pdf），檔案中以專業期刊上所要求的格式，以表格呈現分析結果，並加上適當的分析及討論。

#### 15. [5月30日] 期刊經驗研究成果的重製

#### 16. [6月6日] 進階課題導讀

- 第四次作業繳交

#### 17. [6月13日] 獨立進行研究與報告寫作

#### 18. [6月20日] 課程總結 & 研究成果分享

### 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
  - Fox, J. (2002). *An R and S-Plus companion to applied regression*. Sage Publications.
  - 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

- 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. Teator, 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)