# 國立中山大學政治學研究所
## 政治學量化資料分析

Quantitative Data Analysis in Politcal Science
Fall 2008 (971)

December 24, 2008

Instructor: 劉正山 助理教授
Time: Thursday 2-5 pm
Classroom: 社 3007
Office: 社 4041
Course Website: http://www2.nsysu.edu.tw/politics/liu (updated syllabi will be posted here)

Office Hours: Tue. & Wed. 1-3 pm or by appointment
E-mail: csliu@faculty.nsysu.edu.tw
Phone Number: 07-5252000 ext. 5555

## Course Description

　　近年政治科學發展很重要的趨勢之一，是將統計學和計量經濟學應用到研究議題的發掘和理論的發展上。量化資料分析並不是研究政治現象唯一途徑，但是對於有志於從事政治現象的研究者而言，它是個學門內及跨學門的共通語言，更是探索經驗政治現象與發展理論的基礎能力。因此，本課程將以資料分析的實作爲主、以「基礎統計學」、「多變量資料分析」及「計量經濟學」中的基本觀念爲輔，幫助同學（1）學會使用"R"這個全球通用的開放原始碼（open-source）程式語言進行初級（卡方）和中級（迴歸）的資料分析，以及（2）正確解讀以資料分析爲主要研究方法的論文。

　　這門課不要求修課的同學具有基礎統計的背景，但以本課程的目標以及授課的進度來看，沒有相關背景的同學必須在開學前三週的「複習暖身」階段自行補強對統計學的認識和「政治學研究方法」課程中曾經強調過的理論、模型（模式）和假設之間的關係。此外，無論你有沒有應用統計的背景，你必須好好面對以英文爲主的教材及軟體使用手冊。無論你喜不喜歡使用電腦，你必須面對一段"no pain, no gain"的學習過程。同學必須每次上課帶自己的筆電來課堂。所有的課堂練習及回家作業都必須在自己的電腦上完成。這門課也無暇處理政治學知識論和方法論的問題，因此，學習態度上無法自我調整或是心態上排斥量化資料分析的同學請勿勉強選修。

　　這門課可視爲本所「政治學研究方法」的延伸，也是「比較政治理論」、「國際關係」和「經驗政治理論」等課程的輔助。對於有興趣以量化途徑來寫作碩博士論文的同學來說，這門課則爲必選。

## Course Texts and Readings

### Required Texts

　　I used to ask students to buy text books through the university book store. Unsatified with their service, this year I decided to let students to determine how to buy the books. Students must buy original copies of the texts or used books; photocopied books are prohibited. The comprehensiveness of materials covered in this class is limited, so students are encouraged to search for additional text to assist self-learning.

- [V] Verzani, J. (2005). Using R for Introductory Statistics. Chapman & Hall/CRC.

    - You can find full text of the book (but not printable and downloadable) by typing the book name in google books http://books.google.com.tw/books, where you can buy this book through a link to CRC (free shipping!).
    - The official website for this book http://wiener.math.csi.cuny.edu/UsingR/index.html.

- [LC] 林惠玲、陳正倉。2004。基礎統計學。

    - I use this book as a guideline to help the class quickly pick up basic concepts of statistics and modeling. You are encouraged use this book as a supplement to Verzai's book.

- [LCv2] 林惠玲、陳正倉。2008。基礎統計學（二版）。

– This is the new version of the text, I will use this book next book. In this syllabus, chapters corresponding to [LC] are marked.

## Supplemental Texts (<u>not required to buy</u>)

This is a list of some good books for your reference.

- Some books that make more sense of statistics for you:

  – 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。

- Some books for advanced learners:

  – 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。
  – 王濟川、郭志剛。2005。Logistic迴歸模型：方法及應用。台北：五南。
  – Long, J. Scott. 1997. Regression Models for Categorical and Lmited 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。）
  – Harrell, F. E. (2001). Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. Springer.

- Some books that provide good datasets:

  – Lewis-Beck, M.S. (1995). Data Analysis: An Introduction. Thousand Oaks, CA: Sage Publications.
  – Romer, D.; Jamieson, K. H. & Winneg, K. K. A. C.(Eds.) (2006). Capturing Campaign Dynamics, 2000 And 2004: The National Annenberg Election Survey. Philadelphia: University of Pennsylvania.

- Other introductory books to R and statistics:

  – Dalgaard, P. (2002). Introductory Statistics with R. New York : Springer.
  – Fox, J. (2002). An R and S-Plus companion to applied regression. Sage Publications.
  – 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`.

# About Using R

R (`http://www.r-project.org/index.html`) is a statistical programming language that is widely acknolwdged and used in academics (particularly statistics, biology, ecology, and political science) around the world. Researchers using R can do most of statistical works that other commercial packages–like SPSS, SAS, STATA, and S-plus–can do without any financial cost. Thank communities of scholars and statisticians that continue to develope R, users of R have a great level of freedom.

The difficulty level of learning R is higher than SPSS and a bit higher than STATA, but about the same as SAS and S-Plus. It is mainly because R user interface is not poin-and-click oriented but requires users to be familiar with its language syntax. This barrier has been gradually lowered down; thanks John Fox and many others' effort to improve R-commander (Rcmdr) package that gives R a handy graphical user interface (GUI).

Recommending learning and using R, I don't mean that other commercial packages are worthless. SPSS, for example, is strong in its data editing and storage functions. My point is that students should not buy pirated version of those commercial packages while good alternatives are available. Since R is still a tool new to students of the discipline in Taiwan, there is few Chinese textbooks available to beginners. So I don't prevent students from choosing a commercial package with which they have been familar. In that case, I will recommend using STATA and S-Plus.

## Grading Policy

- 4 Homeworks/Assignments (20%, 20%, 20%, 30%): the assignments need to be done by YOURSELF, cooperation or any kind of team work is not allowed. If you have questions when you work alone, it is ok to consult your classmates. You are encouraged to search for online documents or use my office hours.

- Class participation and attendance (10%).

# Course Schedule

## Part I: Review and Preparation

### 1. [Sept. 18] Introduction

- In Class:

  - explain the syllabus and the purpose of this class

- After Class:

  - use the following week to get the text books, get your laptop ready, install R & Emacs, and read through the instructory pages in a R reference website site constructed by NSYSU library in the summer of 2008 `http://140.117.120.59:88/moodleservice/`.

### 2. [Sept. 25] Warm Up for the Class

- Note: class time will be start at 10:00

- In Class:

  - review the relationships between theory, model, and hypotheses.

- After Class:

  - use Emacs to a txt file of statement of purpose (breifly stating who you are and why you wanted to take this class). E-mail the txt file to the instructor.

### 3. [Oct. 2] Fundamental Concepts for Statistical Analysis

- In Class:

  - talk about causal relationships and measurement issues
  - demonstrate how to work with Emacs, ESS, and R.

- Supplemental:

  - 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.
  - Box, G. E. P. (1979). Some problems of statistics and everyday life. Journal of the American Statistical Association, 74 (365), 1-4.
  - Holland, P. W. (1986). Statistics and Causal Inference. Journal of the American Statistical Association, 81 (396), 945-960.

- After Class:

  - read [V] Appendix B and see if you can install the "Rcmdr" package by yourself.

4. [Oct. 9] Data Formats

- Before Class:

  - [V] Chapter 1;
  - [LC] Chapter 2

- In Class:

  - Import data, export data, save datasets, and create your own datasets.
  - Discuss the structure of three types of datasets: survey, IR, and comparative politics.

- In Class:

  - First Assignment Out.

## Part II: Foundamental Data Analysis with R

### 5. [Oct. 16] Multivariate Datas

- Before Class:

  - [V] 4.1, 4.2, & 4.5

### 6. [Oct. 23] Populations, Sampling, and Sampling Distribution

- Before Class

  - [V] Chapter 5;
  - [LC] Chapters 3 & 9; [LCv2] Chapter 10

- First Assignment Due.

### 7. [Oct. 30] Probability and Distributions

- Before Class

  - [LC] Chapters 6, 7, & 8; [LCv2] Chapters 7, 8 & 9
  - pay close attention to features of normal distribution and Z-score

- Second Assignment Out.

### 8. [Nov. 6] Uni-variate Data Analysis: Variables, Concepts, and Distributions

- Before Class:

  - [V] 2.1- 2.3;
  - [LC] Chapters 4 & 5; [LCv2] Chapters 4 & 5

### 9. [Nov. 13] How to Present Your Tables Professionally

- Before Class:

  - Klass, Gary M. 2008. Just Plain Data Analysis: Finding, Presenting, and Interpreting Social Science Data.Lanham, MD: Rowman & Littlefield. [Chapters 2 and 3].
  - Bring a journal article on political behavior using tables that you think present idea well. You will need to lead the discuss of the chosen article.

### 10. [Nov. 20] Bi-variate Data Analysis: Simple Comparisons

- Before Class:

  - [V] 3.1, 3.3, & 9.2
  - [LC] Chapters 14; [LCv2] Chapter 15

**11. [Nov. 27] Multivariate Data Analysis**

- Second Assignment due

- Before Class:

    - [V] 4.1, 4.2, and Chapter 9;

**12. & 13 [Dec. 4 & 11] Statistical Inference: Confidence Intervals & Significance Tests**

- Confidence Intervals:

    - [V] Chapter 7 (skip 7.4 and 7.6);
    - [LC] Chapters 10 & 11; [LCv2] Chapters 10 & 11

- Significance Tests:

    - [V] Chapter 8;
    - [LC] Chapters 11 & 12

- Supplemental readings about "margin of error":Margin of Sampling Error `http://www.aapor.org/marginofsamplingerror` & Margin of Error: When Used by Pollsters, Is Widely Misunderstood and Confuses Most People `http://www.harrisinteractive.com/harris_poll/index.asp?PID=832`.

## Part III: Advanced Data Analysis with R

**14. [Dec. 18] Linear Regression**

- 3rd assignment out.

- Before Class

    - [V] 3.4, 10.1, 10.2, 10.3
    - [LC] Chapter 13; [LCv2] Chapter 14

- Supplemental:

    - Kam, C. D. & Franzese, R. J. (2007). Modeling and Interpreting Interactive Hypotheses in Regression Analysis. Ann Arbor: University of Michigan Press. – Read Chapter3.

**15. [Dec. 25] Logistic Regression (I): Basics**

- 3rd assignment due & 4th assignment out

- Before Class

    - [V] 12.1
    - supplemental readings

**16. [Jan. 1, 2009] No Class.**

**17. [Jan. 8] Logistic Regression (II): Interpretation and Concerns**

- Before Class

    - [V] 12.1
    - supplemental readings

**18. [Jan. 15] Student Presentation & Semester Conclusion**

- 4th assignment due.